

# IMPZ: a new photometric redshift code for galaxies and quasars

T. S. R. Babbedge,<sup>1\*</sup> M. Rowan-Robinson,<sup>1</sup> E. Gonzalez-Solares,<sup>2</sup> M. Polletta,<sup>3</sup>  
S. Berta,<sup>4</sup> I. Pérez-Fournon,<sup>5</sup> S. Oliver,<sup>6</sup> D. M. Salaman,<sup>6</sup> M. Irwin<sup>2</sup> and S. J.  
Weatherley<sup>1</sup>

<sup>1</sup>*Astrophysics Group, Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2BW*

<sup>2</sup>*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA*

<sup>3</sup>*Infrared Processing and Analysis Center, California Institute of Technology, 100–22, Pasadena, CA 91125, USA*

<sup>4</sup>*Dipartimento di Astronomia, Vicolo dell'Osservatorio 5, I-35122 Padua, Italy*

<sup>5</sup>*Instituto de Astrofísica de Canarias, 38200 La Laguna, Tenerife, Spain*

<sup>6</sup>*Astronomy Centre, CPES, University of Sussex, Falmer, Brighton BN1 9QJ*

Accepted 2004 June 7. Received 2004 April 26; in original form 2003 November 28

## ABSTRACT

We present a combined galaxy–quasar approach to template-fitting photometric redshift techniques and show the method to be a powerful one.

The code (IMPZ) is presented, developed and applied to two spectroscopic redshift catalogues, namely the Isaac Newton Telescope Wide-Angle Survey ELAIS N1 and N2 fields and the *Chandra* Deep Field North. In particular, optical size information is used to improve the redshift determination. The success of the code is shown to be very good, with  $\Delta z/(1+z)$  constrained to within 0.1 for 92 per cent of the galaxies in our sample.

The extension of template fitting to quasars is found to be reasonable, with  $\Delta z/(1+z)$  constrained to within 0.25 for 68 per cent of the quasars in our sample. Various template extensions into the far-ultraviolet are also tested.

**Key words:** galaxies: evolution – galaxies: photometry – quasars: general – cosmology: observations.

## 1 INTRODUCTION

Photometric redshifts are a powerful statistical tool for studies of the evolutionary properties of galaxies, in particular of faint galaxies, for which spectroscopic data are hard or impossible to obtain.

Photometric redshifts are faster to measure than spectroscopic redshifts and can be applied to much fainter magnitudes since the bin sizes are larger ( $\approx 1000$  Å versus 1–2 Å). There is, however, a trade-off with redshift precision – Hogg et al. (1998) found that photometric redshifts can be predicted with an accuracy of  $\Delta z \approx 0.1$  (0.3) in  $z$  for 66 per cent (99 per cent) of the sources examined.

The bulk of the photometric redshift identification is carried out using the broad-band continuum shape and presence and/or absence of spectral breaks (like the 4000-Å Balmer break) or the onset of the Ly $\alpha$  forest and Lyman limit, which enter optical wavebands at high redshift [the Ly $\alpha$  forest effect (e.g. Steidel & Sargent 1987; Madau 1995)]. Although methods based on training sets such as polynomial fitting (e.g. Wang, Bahcall & Turner 1998) and artificial neural networks (e.g. Tagliaferri et al. 2002; Ball et al. 2004) have had some success in determining photometric redshifts, we decided to utilize the template-fitting procedure (but with an element of training in that the templates have been adapted to improve fits).

This is because there is a relative paucity of galaxies with both spectroscopic redshifts and photometry in the bandpasses used in the catalogues under study here, particularly at higher redshifts, making the construction of a training set unfeasible. Empirical methods are also hard to apply outside of the boundaries in which they were defined – such as the redshift distribution of the training set or the photometric bands use.

In the template-fitting method (e.g. Sawicki, Lin & Yee 1997; Giallongo et al. 1998), the observed galaxy fluxes,  $f_i^{\text{obs}}$  in the  $i$ th band, are compared to a library of reference fluxes,  $f_i^{\text{templ}}(z, T)$ , where  $T$  is a set of parameters that account for the template galaxy's morphological type, age, metallicity and dust. We then fitted the observed fluxes to the library fluxes using  $\chi^2$  minimization. As well as deriving redshifts, the procedure produces information on spectral (template) type, although this is less robust as a result of the degeneracies in the parameter space (see Section 6.3.1). A way of breaking these degeneracies is to use Bayesian probability (e.g. Jaynes 2000) to weight the solutions based on a prior knowledge of the expected population distributions. Application of Bayesian methods to photometric redshifts has been presented in, for example, Kodama, Bell & Bower (1999) and Benítez (2000). Usually such applications use priors such as the expected redshift distribution of the sample, but this naturally suppresses unbiased information on the true redshift distribution. Here Bayesian methods are implemented using absolute magnitude limits and extinction distributions, which

\*E-mail: tsb1@imperial.ac.uk

it is hoped retain the power of Bayesian methods without unduly influencing the underlying science (see Section 2.2).

One important refinement is the consideration of extinction for galaxies. Madau (1999) used a single dust absorption correction of  $A_{1500} = 1.2$  mag, except for galaxies in the redshift range 0.75–1.75, where the equivalent extinction at 2800 Å was used. Galaxy evolution models such as those of Le Borgne & Rocca-Volmerange (2002) include evolution of dust extinction with time. In order to allow for variation in extinction from galaxy to galaxy, extinction needs to be solved as an additional free parameter to redshift. In Steidel et al. (1999), dust absorption was corrected for by assuming that colour deviations in their sample galaxies were entirely due to dust absorption, based on the empirical relation between far-infrared (far-IR) emission and the observed ultraviolet (UV) spectrum slope, as derived by Meurer, Heckman & Calzetti (1999). More recently, Thompson, Weymann & Storrie-Lombardi (2001) used spectral energy distribution (SED) template-fitting photometric redshift techniques in the deep NICMOS northern *Hubble Deep Field* (HDF), fitting extinction as a parameter ranging from  $E(B - V) = 0$  to 1.0. The study of Bolzonella, Miralles & Pelló (2000) found that the inclusion of  $A_V$  as a free parameter caused significant increases in aliasing. In a similar technique developed in (Rowan-Robinson 2003b, hereafter RR03), these aliasing problems were reduced by setting several  $A_V$  priors (see Section 2.2).

In this paper the SED template fitting set out in RR03 is refined and applied to two spectroscopic redshift galaxy samples from the European Large-Area *ISO* Survey (ELAIS; Oliver et al. 2000) N1 and N2 fields of the Isaac Newton Telescope Wide-Angle Survey (INT WAS; McMahon et al. 2001), a part of the INT Wide-Field Survey (INT WFS), and also the *Chandra* Deep Field North (CDFN) for validation purposes. The effect of non-zero  $A_V$ , different SED templates and various template extensions into the far-UV are explored, as is the inclusion of several different quasar templates and the applicability of template-fitting techniques to quasar-like sources. In Section 2 the photometric redshift technique is set out. In Section 3 the various templates and extensions to the UV are discussed. The IMPZ code is then applied to two spectroscopic redshift galaxy samples in Section 4. The results of this validation are given in Section 5. Error analysis is discussed in Section 6. Discussions and conclusions are presented in Sections 7 and 8.

The application of the IMPZ code to the entire recalibrated ELAIS N1 and N2 fields from the INT WAS and investigations into the evolution of extinction and star formation rates (SFR) will be presented in a companion to this paper (Babbedge et al., in preparation).

Note that for these investigations the flat,  $\Omega_\Lambda = 0.7$  cosmological model with  $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$  is used.

## 2 METHOD

### 2.1 $\chi^2$ analysis

The template-fitting procedure is as follows: The observed galaxy magnitudes are converted for each  $i$ th photometric band into an apparent flux,  $f_i^{\text{obs}}$ . Equivalently, this reconstructs the SED of the target galaxy at a very low spectral resolution by sampling the luminosity at the effective wavelength of each photometric band. The observed fluxes can then be compared to the template fluxes,  $f_i^{\text{templ}}(z, T)$ , computing the reduced  $\chi^2$ ,  $\chi_{\text{red}}^2$  as

$$\chi_{\text{red}}^2 = \frac{1}{D} \sum_{i=1}^N \frac{[f_i^{\text{obs}} - s f_i^{\text{templ}}(z, T)]^2}{\sigma_i^2}. \quad (1)$$

Here  $N$  is the number of photometric bands;  $D$  is the number of degrees of freedom;  $\sigma_i$  is the observational uncertainty in the  $i$ th band (hence the solution is weighted by the flux errors, but to avoid excessively high weighting by very high signal-to-noise ratio observations, a minimum flux error is set, typically 0.5 per cent); and  $s$  is a normalization factor to minimize  $\chi^2$  for each template:

$$s = \frac{\sum_{i=1}^N f_i^{\text{obs}} f_i^{\text{templ}}(z, T) / \sigma_i^2}{\sum_{i=1}^N f_i^{\text{templ}}(z, T)^2 / \sigma_i^2}. \quad (2)$$

Note that if there is a detection in just one band, then fitting is not attempted.

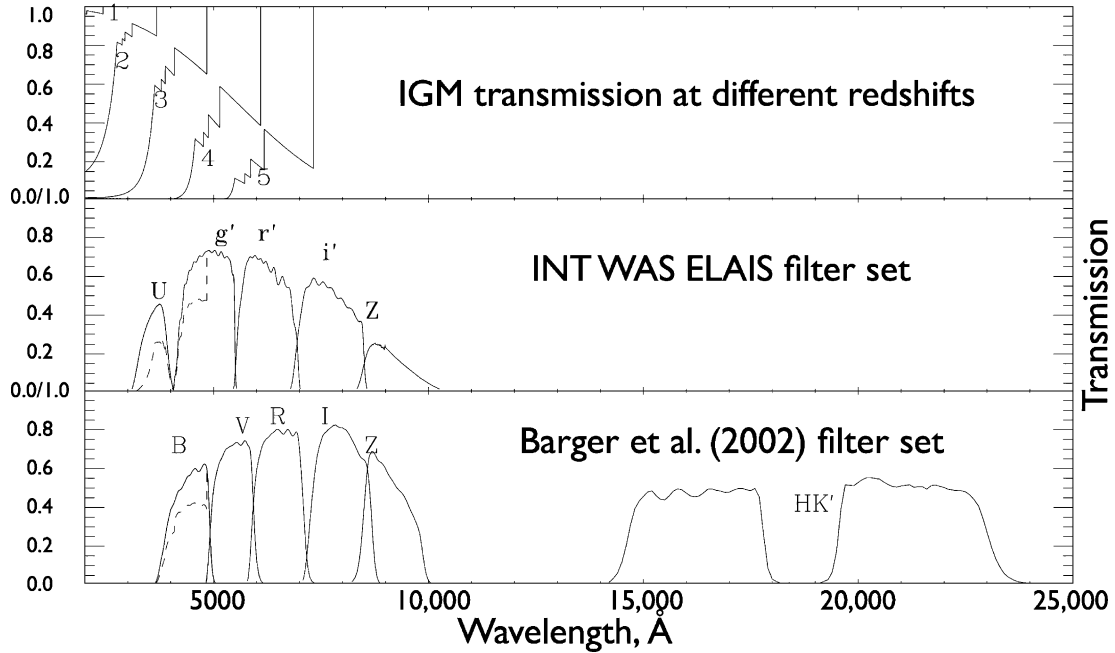
### 2.2 The IMPZ code

This code builds on the technique presented in RR03, extending it to include the correct treatment of charge-coupled device (CCD) response, filter transmission characteristics and the statistical effect of intergalactic medium (IGM) absorption ( $\text{Ly}\alpha$ ,  $\text{Ly}\beta$ ,  $\text{Ly}\gamma$ ,  $\text{Ly}\delta$  and Lyman continuum), as set out in Madau et al. (1996) (see Fig. 1). Correct treatment of the effects of the IGM is needed because it is possible to misinterpret the  $\text{Ly}\alpha$  forest effect as the intrinsic Lyman break and because, particularly at high redshifts, it will imprint its own recognizable feature on to the SED. Although the treatment of IGM absorption is only based on the average accumulated absorption, the rms fluctuations away from the mean can be expected to be small once integrated through a broad bandpass (Press, Rybicki & Schneider 1993). At high redshift, Massarotti, Iovino & Buzzoni (2001) have shown that correct treatment of internal dust reddening (the interstellar medium) and IGM attenuation are the main factors in photometric redshift success. The effect of internal dust reddening for each galaxy is already incorporated in the templates (see Table 1) and alterable via fitting for  $A_V$ , using the reddening curve of Savage & Mathis (1979). Observed fluxes are compared to template fluxes for  $0.01 \leq \log_{10}(1+z) \leq 0.90$ , equivalent to  $0.02 < z \leq 6.94$ .

The following parameters and cuts were used:

(i) In order to reject unphysical fits, only those that give absolute  $B$ -band AB magnitudes ( $M_B$ , 4400 Å) in the range  $[-22.5 - 2 \log_{10}(1+z)] < M_B < -13.5$  are considered for the six galaxy templates. Having an upper envelope dependent on redshift was found to be the best way to suppress luminous outliers at low redshift whilst allowing more luminous galaxies at higher redshifts, and this luminosity–redshift dependence agrees with the natural consequence of known strong luminosity evolution for galaxies (e.g. Lin et al. 1999). For the active galactic nuclei (AGN) templates, a range from  $-27 < M_B < -17.5$  was allowed. These restrictions cut out excessive numbers of aliases at the minimum and maximum redshift. A number of different limits have been investigated, but these were found to be the most effective. Originally, limits were applied to the  $I$  band, but this failed to constrain the UV luminosity. Shifting the limits to the  $B$  band allows both the young and old star components of the SED to be constrained for the redshifts considered.

(ii) Sources are defined as *stellar* or *non-stellar*, where this is a purely morphological property differentiating between point-like and more extended sources. Those defined as *stellar* will have AGN templates considered in addition to the galaxy templates, whereas *non-stellar* sources will only be fitted by galaxy templates. This reduces the increased degeneracy introduced by including AGN templates. The procedure for splitting sources into *stellar/non-stellar* subgroups is different for the INT WAS ELAIS spectroscopic



**Figure 1.** IGM transmission effect at redshift 1, 2, 3, 4 and 5 on the observed flux from a source. Below are shown the INT WAS filters (solid lines), combined with the CCD response of the WFS camera, and also the filters (solid lines) used in the Barger et al. (2002) CDFN catalogue, combined with the relevant CCD (Suprime Cam and QUIRC) responses. The additional effect of including IGM absorption is illustrated for the redshift 3 case by overplotting the filter response when combined with the redshift 3 IGM transmission curve. It can be seen to affect *U* and *g'* in the INT WAS filter set, and the *B* filter in the CDFN filter set (overplotted as dashed lines).

sample and the CDFN sample, and is described for each in Sections 4.1 and 4.2, respectively.

(iii) Sources that were saturated were removed since their photometry is poor. This reduces  $z_{\text{phot}}$  outliers.

(iv) A prior expectation that a *stellar* source is more likely to be a quasi-stellar object (QSO) is introduced by minimizing  $\chi^2_{\text{red}} + \alpha F(T)$  rather than  $\chi^2$  ( $\alpha = 4$  here) for *stellar* sources. Here  $F(T)$  is a delta function such that  $F(T) = 0$  if the template  $T$  is an AGN template, and  $F(T) = 1$  if the template is a galaxy template. Essentially it is a prior to prefer AGN solutions due to the morphology information contained in the class flags – a weak Bayesian (e.g. Jaynes 2000) formulation – and was reached based on the results in Sections 5.1.4 and 5.2.3.

For the ‘free’  $A_v$  fitting option, the following restrictions were used:

(v) For the elliptical and AGN templates,  $A_v$  can take the value 0 only. The reason that  $A_v$  is set to zero for AGN is that, since AGN are essentially a power law, the additional inclusion of  $A_v$  gives too much freedom to the shape of the AGN template, and it was found that resulting degeneracies reduce the effectiveness of the photometric redshift technique.

(vi) For other templates,  $A_v$  can take the values 0.0 to 1.0, in steps of 0.1. The maximum  $A_v$  was chosen to be approximately twice that of the typical  $A_v$  of galaxies at  $z \approx 4$  found by Steidel et al. (1999), who derived a typical  $E(B - V)$  of 0.15. Note that the templates already include some extinction (see Table 1) so that the  $A_v$  of the solution is technically the difference between the actual value and the template’s inherent value.

(vii) No solution for  $A_v$  is sought if the reduced  $\chi^2$ ,  $\chi^2_{\text{red}}$ , of the  $A_v = 0$  solution is  $< 1$ , or if there are less than four bands.

(viii) A prior expectation that the probability of a given value of  $A_v$  declines as  $|A_v|$  moves away from 0 is introduced by minimizing

$\chi^2_{\text{red}} + \beta A_v^2$  rather than  $\chi^2$  ( $\beta = 2$  here). This can again be viewed as a weak implementation of Bayesian methods.

### 3 TEMPLATE SEDS

The choice of how many templates to use in fitting is a crucial one. The choice of too many leaves the code with too much freedom, leading to large numbers of aliases and degeneracy. Similarly, too few and the code will be unable to find accurate redshifts for real objects, something Bolzonella et al. (2000) termed ‘catastrophic’ failures.

As in Mobasher et al. (1996) and RR03, the code uses six galaxy templates: E, Sab, Sbc, Scd, Sdm and starburst (Sbrst) galaxies. These six templates (or similar versions) have been found to provide a good low-resolution representation of observed galaxy SEDs (Mobasher et al. 1996; RR03), and are used in preference to a large array of evolving galaxy templates, as generated by evolutionary codes such as that of Bruzual & Charlot (1993), since too much freedom is then available in fitting. The empirical templates used in RR03 were based on observations (starburst template adapted from Calzetti & Kinney 1992) and on observations and colour synthesis (the remainders adapted from Yoshii & Takahara 1988). In a method reminiscent of empirical techniques, these templates were adapted to improve photometric redshift results.

The templates used in this work were generated by reproducing these original templates via spectrophotometric synthesis, in order to strengthen their physical basis. The original templates were convolved with filter transmission curves in order to create virtual data points. These data points were then fitted using a code that combines a given number of simple stellar populations (SSPs), each weighted by a different SFR and extinguished by a different amount of dust. The procedure is based on the synthesis code of Poggianti, Bressan

**Table 1.** The six galaxy templates and the SSPs that were used to create them. The SFR is given in  $M_{\odot} \text{ yr}^{-1}$  and normalized so that each template has a total mass of  $10^{10} M_{\odot}$ .

SSP		SED template									
Age (yr)	$\Delta t$ (yr)	Elliptical (E1)		Sab		Sbc		Scd		Sdm	
		SFR	$E(B - V)$	SFR	$E(B - V)$	SFR	$E(B - V)$	SFR	$E(B - V)$	SFR	$E(B - V)$
1	$10^6$	—	—	—	—	—	—	—	—	—	—
2	$3 \times 10^6$	—	—	0.100	0.04	—	—	—	—	—	—
3	$8 \times 10^6$	—	—	—	—	—	—	—	—	41.17	0.9465
4	$10^7$	0.099	—	—	—	1.461	0.18	—	—	35.67	0.024
5	$5 \times 10^7$	—	—	0.332	0.015	—	—	2.576	0.03475	—	—
6	$1.25 \times 10^8$	—	—	—	—	—	—	—	—	0.001	—
7	$3 \times 10^8$	—	—	0.484	0.12	1.647	—	1.692	0.12	13.54	0.004
8	$5 \times 10^8$	—	—	—	—	0.896	0.003	2.80	0.032	13.62	4.685
9	$10^9$	0.396	—	1.074	0.017	1.509	0.137	0.771	0.1838	9.995	—
10	$1.2 \times 10^{10}$	0.950	—	0.854	0.134	0.747	0.186	0.756	0.138	11.58	0.0616
$A_v(\text{young} - \text{first three SSPs})$		0.0	0.0	0.12205	0.0	0.0	0.0	0.0	0.0	0.0031	1.38
$A_v(\text{total})$		0.0	0.0	0.2344	0.0	0.3308	0.0	0.271	0.0	0.4303	0.736

& Franceschini (2001) and has previously been applied to spectra in Berta et al. (2003). The code minimizes the  $\chi^2$  obtained from comparing the new template to the data points. Minimization is based on the Adaptive Simulated Annealing algorithm. Details on this algorithm and on the fitting technique are given in Berta et al. (2004). The templates are plotted in Fig. 2, and have been extended further into the UV regime as set out in Section 3.3.

### 3.1 SSP populations

The spectra of the SSPs have been computed with a Salpeter initial mass function (IMF) between 0.15 and 120 solar masses, adopting the Pickles (1998) spectral atlas and extending its atmospheres outside its original range of wavelengths with Kurucz (1993) models from 1000 to 50 000 Å, as described in Bressan, Granato & Silva (1998). Nebular emission is added by means of case B H II region models computed through the ionization code CLOUDY of Ferland et al. (1998). The adopted metallicity is solar.

When generating the new templates, each SSP is weighted by a different SFR. For each SSP a uniform screen attenuation is adopted, using the standard extinction law [ $R_v = A_v/E(B - V) = 3.1$  (Cardelli, Clayton & Mathis 1989)] and adopting a different  $E(B - V)$ .

With 10 SSPs, we have a total of 20 free parameters, i.e. SFR and  $E(B - V)$  for each SSP, but the code automatically discards those populations which contribute less than 1 per cent to the total spectrum, at each wavelength. As a result, each template is constructed from only a few SSPs. The  $E(B - V)$  values for the oldest populations (ninth and tenth) are constrained to be less than 0.2 [the characteristic extinction of the older quiescent stellar population in a sample of nearby galaxies is  $E(B - V) \approx 0.1$  (Rowan-Robinson 2003a)]. The total  $E(B - V)$  for each template is obtained by comparing the non-extincted final spectrum ( $S_n$ ) to the extincted version ( $S_e$ ), since in the  $V$  band the following relation holds:  $S_e = 10^{-0.4A_v} S_n$ . For details of the populations and their contribution to each template, see Table 1.

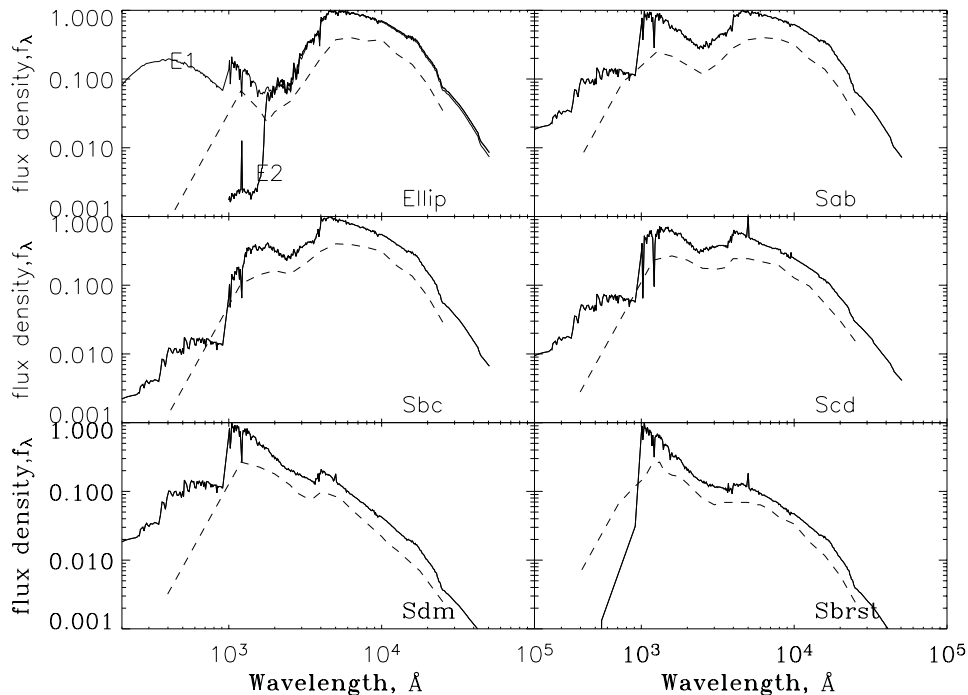
Two fits were generated for the elliptical template: E1 and E2. E1 fits the small UV bump present in ellipticals at approximately 1000–2000 Å, a feature that is due to emission from planetary nebulae (Yoshii & Takahara 1988). In order to prevent this UV bump being fitted by young stellar populations, the SFR of the three youngest SSPs was set to zero. E2 consists only of the two oldest SSPs and fails to fit this UV bump. For this reason, E2 was not used as a template, but is plotted in Fig. 2 for interest.

The template extension into the UV is described in Section 3.3.

### 3.2 AGN templates

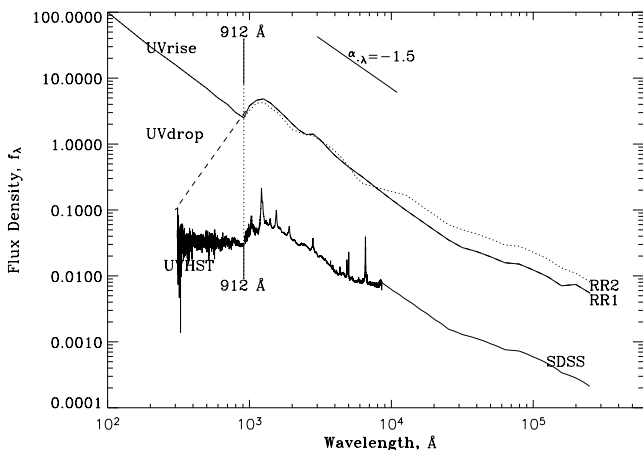
As well as galaxy templates, the inclusion of a number of different AGN templates has been investigated (see Fig. 3) to allow the IMPZ code to identify quasar-type objects as well as normal galaxies. This is of particular interest for application of IMPZ to the entire ELAIS N1 and N2 fields of the INT WAS in Babbedge et al. (in preparation) since many ELAIS sources are expected to be AGN (Oliver et al. 2000). Fitting with AGN templates is only carried out for sources that have been defined as *stellar*, as described in Section 2.2.

The last decade has seen a large rise in the number of optically selected high-redshift quasars, and the existence of large samples of quasars, e.g. the Sloan Digital Sky Survey (SDSS, York et al. 2000), means that the derivation of photometric redshifts for quasars is gaining popularity as well as reliability. For galaxies the technique



**Figure 2.** The six galaxy templates used. Dashed lines show the original RR03 templates (offset for clarity); solid lines show the SSP generated versions, along with extension into the far-UV (sub-1000 Å) as set out in Section 3.3. For the elliptical template, two SSP generated fits are shown, which diverge below around 2000 Å. Line E1 fits the UV bump that is due to planetary nebulae, whereas line E2 does not. This is discussed in Section 3.1.

relies on the identification of continuum features such as the 4000-Å break – see Table 2 for examples. For a featureless spectrum, a photometric redshift is far harder to determine, if at all. The majority of quasars can be characterized in the UV–optical region as a featureless continuum. Overlaid on this continuum, however, are a series of (mostly) broad emission-line features that contain a significant amount of flux (Francis et al. 1991; Richards et al. 2001a). Also, at higher redshifts the result of Ly $\alpha$  forest absorption will imprint an



**Figure 3.** The various AGN templates that were investigated. The solid line labelled ‘SDSS’ is the mean SDSS quasar spectra (Vanden Berk et al. 2001), shown here with the Zheng et al. (1997) UV behaviour, ‘UVHST’, and extended into the IR as set out in Section 3.2. The solid line labelled ‘RR1’ and the dotted line ‘RR2’ are the empirical AGN templates based on Rowan-Robinson (1995), shown with either a drop-off in the UV (‘UVdrop’, dashed) or a rise in the UV (‘UVrise’, solid). The 912-Å Lyman limit has been indicated, as is the slope of a power-law continuum with  $\alpha_\lambda = -1.5$ . Discussion of these AGN is covered in Section 3.2, with UV treatment discussed in Section 3.3.

additional redshift-dependent feature on to the continuum. Empirical redshift–colour relationships have been applied to quasars, but such polynomial fitting is limited because of the nature of quasar colours, which can change rapidly across a small redshift range as an emission line passes in to and out of a passband, or remain constant with redshift due to the featureless continuum – a relationship that is poorly reproduced by polynomial functions that vary slowly with redshift. An extension of this technique which implements a nearest-neighbour (NN) estimator with the reference points derived from colours averaged over redshift bins was presented in Richards et al. (2001b), with around 70 per cent of predicted redshifts matching reasonably well with the spectroscopic values ( $\Delta z < 0.1$  for 55 per cent and  $\Delta z < 0.2$  for 70 per cent of their sample).

The alternative approach is a template-fitting procedure as used in several quasar studies (e.g. Hatziminaoglou, Mathez & Pelló 2000) and adopted in this work. This approach requires a template or templates that can cover the observed range of spectral types, across a large enough wavelength span to be applicable for low- and high-redshift objects. Unlike galaxies, quasars have SEDs with similar power-law continua, so the use of the mean spectrum of a sample of quasars is feasible. Budavári et al. (2001) used the SDSS composite spectrum of Vanden Berk et al. (2001), resulting in a slightly greater rms than that found in the empirical NN method of Richards et al. (2001b). This suggests that the use of a single quasar template is not sufficient – perhaps one corresponding to broad absorption lines or Fe II is also needed. A similar study by Wu, Zhang & Zhou (2004) with the SDSS composite achieved  $\Delta z < 0.1$  for 47 per cent and  $\Delta z < 0.2$  for 68 per cent of their sample. Budavári et al. (2001) went on to reconstruct four discrete templates in an iterative manner, gaining better results than with the empirical NN method. However, this is beyond the scope of this paper and is left as a future direction.

In this study, several AGN templates were constructed. The optical–IR basis of these templates is outlined below, and extension to UV in Section 3.3.

**Table 2.** Common spectral lines and breaks in SEDs, and the redshifts where they enter and leave the filters used in this study, where transition is defined by the points where filter transmission drops below 50 per cent of the maximum filter transmission. Redshifts greater than 7 are not considered.

Filter	Lyman limit (912 Å)		Ly $\alpha$ (1216 Å)		Balmer break (4000 Å)		$\overline{\text{O}}[\text{III}]$ doublet (4983 Å)		H $\alpha$ (6563 Å)	
	enters	leaves	enters	leaves	enters	leaves	enters	leaves	enters	leaves
(a) INT WAS filters										
<i>U</i>	2.5	3.3	1.6	2.2	–	–	–	–	–	–
<i>g'</i>	3.5	4.9	2.4	3.4	0.03	0.35	0	0.08	–	–
<i>r'</i>	5.0	6.6	3.5	4.7	0.38	0.73	0.10	0.38	0.05	0.07
<i>i'</i>	6.7	NA	4.8	6.0	0.75	1.1	0.40	0.71	0.07	0.30
<i>Z</i>	NA	NA	6.0	6.9	1.1	1.4	0.71	0.93	0.30	0.46
(b) Barger et al. (2000) filters										
<i>B</i>	3.3	4.5	2.2	3.1	0	0.25	0	0.003	–	–
<i>V</i>	4.5	5.6	3.1	3.9	0.25	0.50	0.003	0.20	–	–
<i>R</i>	5.6	6.7	3.9	4.8	0.50	0.75	0.20	0.40	0	0.07
<i>I</i>	6.9	NA	4.9	6.1	0.80	1.2	0.44	0.73	0.10	0.31
<i>Z</i>	NA	NA	5.5	NA	1.1	1.5	0.69	1.0	0.28	0.52
<i>HK'</i> , ( <i>H</i> )	NA	NA	NA	NA	2.7	3.5	2.0	2.6	1.2	1.7
<i>HK'</i> , ( <i>K'</i> )	NA	NA	NA	NA	3.9	4.8	2.9	3.6	2.0	2.5

(i) The SDSS median composite quasar spectrum (Vanden Berk et al. 2001) covers a rest-wavelength range from 800 to 8555 Å and is constructed from 2204 quasars spanning  $0.044 \leq z \leq 4.789$ . For wavelengths longward of Ly $\alpha$  to optical wavelengths the continuum is well fitted by a power law,  $f(\lambda) \propto \lambda^{\alpha_\lambda}$ , with a wavelength power-law index,  $\alpha_\lambda$  [ $\alpha_\lambda = -(\alpha_v + 2)$ ], of  $-1.56$ . This is consistent with a number of other works based on optical and/or radio-selected samples that have found power-law continuum indices, as average values from spectra, from photometry, or using composite spectra, of  $-2 \leq \alpha_\lambda \leq -1$  (e.g. Cristiani & Vio 1990; Francis et al. 1991; Francis 1996; Zheng et al. 1997; Natali et al. 1998; Carballo et al. 1999; Brotherton et al. 2001; Schneider et al. 2001). The continuum blueward of the Ly $\alpha$  emission line is heavily absorbed (the median redshift is 1.253) due to Ly $\alpha$  forest absorption. Since this effect is strongly affected by redshift, however, and the median composite uses spectra across a broad range of redshift, little can be drawn from the absorption in that region of the SED. Instead, the UV part of the SED will be treated separately – see Section 3.3. The template has been extended to wavelengths longer than 8555 Å by utilizing the IR part of the average optical quasar spectrum of Rowan-Robinson (1995), slightly modified as set out in Rowan-Robinson et al. (2004). This extends the template out to 25  $\mu\text{m}$  with what is essentially a continuation of the continuum power law. See the ‘SDSS’ line in Fig. 3.

(ii) In addition to the SDSS composite template, two simpler AGN templates were included. These are based on the mean optical quasar spectrum of Rowan-Robinson (1995), spanning 400 Å to 25  $\mu\text{m}$ . For wavelengths longer than Ly $\alpha$  the templates are essentially  $\alpha_\lambda \approx -1.5$  power laws, with slight variations included to take account of observed SEDs of ELAIS AGN (Rowan-Robinson et al. 2004). These two AGN templates are referred to as RR1 and RR2, and are very similar to those used in Rowan-Robinson et al. (2004). They differ from one another at wavelengths longer than 1  $\mu\text{m}$ , where RR2 contains more flux. For wavelengths shorter than the Lyman limit, several UV behaviours were again considered – see Section 3.3. These templates can be seen as ‘RR1’ and ‘RR2’ in Fig. 3.

### 3.2.1 A red quasar

Should a much redder quasar be considered? There is a debate over the existence of a significant population of *red* quasars (e.g. Webster

et al. 1995; Brotherton et al. 2001). Richards et al. (2003) found that roughly 6 per cent of their homogeneously selected sample of 4576 SDSS quasars were red in comparison with even the reddest power-law continuum quasars and were probably dust-reddened. In this work a *red* AGN template has not been included – some tests were done using the  $z = 2.216$  FIRST J013435.7–093102 source from Gregg et al. (2002), which is an extremely dust-reddened lensed object with  $B - K \geq 10$ , but it was found that it did not match any sources in our two spectroscopic redshift catalogues. The inclusion of a red AGN template is, however, expected to be powerful when optical data are combined with upcoming IR data from the *Spitzer* mission, particularly for the *Spitzer* Wide-Area Infrared Extragalactic Survey (SWIRE; Lonsdale et al. 2003), which covers an area large enough (50 deg<sup>2</sup>) to find a significant number of these rare objects. Application to SWIRE data will be considered in a future work (Babbedge, in preparation).

### 3.3 Far-UV treatment

The Milky Way becomes virtually opaque at wavelengths between 100 and 912 Å as a result of absorption by neutral hydrogen. Hence, it is difficult to be certain of the far-UV rest-frame emission of galaxies or quasars. Ground-based observations can only start revealing the sub-1200-Å regime for objects at  $z \gtrsim 2$  because of atmospheric absorption of shorter wavelengths. Escaping our own atmosphere with space-based telescopes has allowed the far-UV and extreme-UV regions to be explored in more detail, but even then the action of the IGM and of neutral hydrogen in our own Galaxy makes the SED determination uncertain. In order to apply a template-fitting technique out to large redshifts, it is important that the templates extend to sufficiently short wavelengths. For galaxies, this is perhaps a more straightforward task, since the emitted electromagnetic energy can be assumed to be due to stars and dust heated by those stars. There exist a number of stellar synthesis codes that, when coupled with spectral evolution, can self-consistently reproduce the emission of galaxies across many orders of magnitude in wavelength. The SSPs adopted by the spectral synthesis code and used to generate the six galaxy templates in this work were not detailed enough below 1000 Å; hence for the far-UV part of these templates the results of other spectral evolution codes has been considered.

Both the isochrone synthesis of Bruzual & Charlot (1993) and the PEGASE code of Fioc & Rocca-Volmerange (1997) show that,

for old stellar populations (older than several Gyr), such as those that characterize ellipticals, there is a rise in the far-UV due to low-mass stars in their post-AGB (asymptotic giant branch) evolution. Furthermore, hot post-AGB stars decrease the amplitude of the 912-Å break once their envelopes have dissipated.

In order to extend the elliptical template into the far-UV, therefore, we used the sub-912-Å part of the spectrum from the elliptical template in HYPERZ from (Bolzonella et al. 2000), who had extended it from the elliptical template of (Coleman, Wu & Weedman 1980, hereafter CWW). The flux was scaled in order to give a slight rise across the 912-Å discontinuity. The four-template set of CWW (elliptical, Sbc, Scd and Im) has been used in many photometric redshift studies (e.g. Gwyn 1999; Benítez 1999; Brodwin, Lilly & Crampton 1999) and has been found to be a robust and reasonably complete template set. The extension into the far-UV carried out by Bolzonella et al. (2000) was by means of Bruzual & Charlot (1993) spectra with parameters (SFR and age) selected to match the observed spectra at zero redshift. They in fact use the IMF of Miller & Scalo (1979); however, this choice has a negligible impact on the final results, as they discuss in their section 4.6.

The Sab, Sbc, Scd and Sdm templates were extended in a similar manner, using the sub-912-Å part of the spectrum (suitably scaled to give a factor of 10 rise in flux across the 912-Å discontinuity) from the Scd template of CWW, extended by Bolzonella et al. (2000). They were all given the same UV behaviour because the difference in UV spectra between these galaxy types is expected to be less than the uncertainty in their actual UV behaviour. Indeed, at redshifts where this region enters the optical filters, the dominant effect is due to IGM absorption.

The starburst template was extended following the results of Bruzual & Charlot (1993) and Fioc & Rocca-Volmerange (1997), which show that within 10 Myr the UV light drops sharply as a result of the evolution of massive stars off the main sequence. Below 912 Å, the starburst was assumed to be optically thick, an assumption that has been verified for nearby starburst galaxies by Leitherer et al. (1995).

It is noted here that, as well as the UV behaviours used above for the six galaxy templates, other forms were tested, such as sharp cut-offs at 912 Å – essentially assuming that the galaxies are optically thick to ionizing radiation below the Lyman limit – or simply taking the flux at 1000 Å and setting this value for sub-1000-Å wavelengths. Such approaches were not found to be as successful.

Determining the exact shape of the UV spectra of AGN is problematic for the same reasons as for galaxies. Additionally, the observed broad continuum feature in the optical–UV, the ‘big blue bump’, is confused by many broad and blended lines, which are thought to be due to fast-moving ionized clouds near the centre of the AGN. Contamination from the host galaxy also has an effect for low-luminosity AGN, particularly in the case of a nuclear starburst. Hence several far-UV trends have been investigated to see how they affect the accuracy of quasar photometric redshifts, as follows.

(i) Although observations of the continuum blueward of the Lyman edge are rare, Zheng et al. (1997) constructed a composite spectrum from 284 *Hubble Space Telescope* (HST) Faint Object Spectrograph (FOS) spectra of 101 quasars at  $z \approx 1$ . Around 90 per cent of the sample were at redshifts less than 1.5, so the region blueward of Ly $\alpha$  could be studied without large effects from the Ly $\alpha$  forest. The shortest-wavelength data, 350–600 Å, were drawn from higher-redshift quasars, for which significant corrections for the Ly $\alpha$  forest and continuum absorption were made. There appears to be a break in the continuum slope at around 1050 Å with  $\alpha_\lambda \approx$

–0.2 in the far-UV. This composite spectrum was used as one possible UV extension to the sub-912-Å SEDs of the ‘SDSS’, ‘RR1’ and ‘RR2’ AGN templates and can be seen in Fig. 3 as the line labelled ‘UVHST’.

(ii) The ‘UVHST’ is approximately a flat continuum. In order to explore the two alternatives, the sub-912-Å SEDs of the ‘SDSS’, ‘RR1’ and ‘RR2’ AGN templates were also modelled as a ‘UVrise’ with  $\alpha_\lambda = -1.5$  and a ‘UVdrop’ with  $\alpha_\lambda = +3$ . These can also be seen in Fig. 3. Recall that IGM absorption is applied to these SEDs depending on redshift.

## 4 SPECTROSCOPIC COMPARISONS

An important stage in the development of any photometric redshift code is to run it on a catalogue of objects for which the redshifts are already known from spectroscopic observations. A version of the IMPZ code has already been used to study the evolution of the UV radiation density, the dust opacity, and hence the star formation history for galaxies in the *Hubble Deep Field* North and South (RR03). As part of this, a comparison was made between the photometric output and the spectroscopic redshifts of 152 HDF-N galaxies. For galaxies with at least four photometric bands ( $U, B, V, I$ ), it was found that the spectroscopic redshifts were successfully matched to an accuracy of around 10 per cent in  $(1+z)$ . Around 2.5 per cent [defined where there was one or more secondary minima with  $\chi^2$  less than 1.0 above the global minimum and  $(1+z)$  differing by more than 20 per cent] of galaxies were found to have a ‘significant’ redshift alias.

The version of IMPZ in RR03 has also been used to generate photometric redshifts for the final band-merged ELAIS catalogue (6.7, 15, 90, 175  $\mu\text{m}$ , and  $U, g', r', i', Z, J, H, K$ , and 20 cm; Rowan-Robinson et al. 2004), fitting both galaxies and quasars by using their  $U, g', r', i', Z, J, H$  and  $K$  data. Again, the photometric redshifts were accurate to around 10 per cent in  $(1+z)$ , with a greater dispersion in  $\Delta z$  for AGN fits.

In order to extend the applicability of IMPZ, this investigation explores IMPZ’s application to two further sets of photometric data with spectroscopic redshifts.

### 4.1 ELAIS spectroscopic redshifts

The spectroscopic data in the ELAIS N1 and N2 fields (Pérez-Fournon et al., in preparation) comprises two samples:

(a) objects observed with the William Herschel Telescope and the fibre-fed WYFFOS/Autofib2 spectrograph at the Observatorio del Roque de Los Muchachos, La Palma, as part of an International Time Project approved by the Comité Científico Internacional of the Observatories of the Instituto de Astrofísica de Canarias for the year 2000 (PI I. Pérez-Fournon); and

(b) objects from the SDSS First Data Release (Abazajian et al. 2003).

These data comprise 172 extragalactic sources, with spectroscopic redshifts ranging from 0.0264 to 2.9426, and detections in none to five of the INT WAS bands. The mean limiting magnitudes (Vega,  $5\sigma$  in 600 s) in each filter across these two areas are 23.40 ( $U$ ), 24.94 ( $g'$ ), 24.04 ( $r'$ ), 23.18 ( $i'$ ) and 21.90 ( $Z$ ). The spectroscopic redshift information also includes a ‘QSO’/‘Galaxy’ flag (26/146 objects). These data are reduced to 163 by removing objects flagged as saturated via the INT WAS class flags (one source), objects that have no WAS detections (four sources), objects with only one detection (three sources), and objects flagged as being contaminated

by a bright star (one source), leaving 157 objects with detections in five bands, and six with detections in four bands, subdivided into 25 ‘QSO’ and 138 ‘Galaxy’ sources. The ‘QSO’/‘Galaxy’ information for INT WAS data enables us to decide on the best combination of the INT WAS class flags for defining possible AGN that should be included in AGN template fitting, and additionally gives a direct measure of how well the galaxy and AGN templates manage to separate out these two populations. Hence a choice can then be made of which AGN templates to use, and what treatment in the UV is most successful. Sources are defined as *stellar* (and so AGN templates are considered in addition to the galaxy templates) based on class flags. The INT WAS sources are classed in each band as follows:  $-1$ , point-like;  $0$ , noise;  $1$ , non-stellar;  $-9$ , saturated;  $-2$ , could be stellar;  $-3$ , might just be stellar. The best bands to define stellarity are  $g'$ ,  $r'$  and  $i'$ , so if the flag is  $-1$  in any of these bands then the source is defined as *stellar*. The choice of this definition of *stellar* was reached after extensive tests with the spectroscopic INT WAS ELAIS catalogue, which found that this identified the most number of actual quasars whilst keeping galaxy contamination to a minimum.

This *stellar* definition splits the sample into 24 *stellar* and 139 *non-stellar* sources.

Galactic extinction in ELAIS N1 and N2 is low (they were after all selected for their low  $100\text{-}\mu\text{m}$  intensity). Both have extinctions of  $E(B - V) = 0.007$ , which using the extinction–wavelength relation in Cardelli et al. (1989) equates to  $A_V \approx 0.022$  for  $R_V = 3.1$ . We can also use the same relation to find the extinction in each of the WAS bandpasses and correct for it (see Table 3). The effect of varying  $R_V$  on the shape of the extinction curve is most apparent at the shorter wavelengths, and, along with uncertainties in the actual form of the average  $R_V$ -dependent extinction law, the accuracy of these corrections is around  $0.002$  in magnitude (less than errors in the actual photometry).

#### 4.2 CDFN spectroscopic redshifts

The Barger et al. (2002) catalogue in the CDFN was also used. This comprises an X-ray-selected catalogue (from the 1 Ms *Chandra* observation of the HDF-N; Brandt et al. 2001) of 169 objects with spectroscopic redshifts and broad-band photometry ( $B$ ,  $V$ ,  $R$ ,  $I$ ,  $z'$  from *Subaru*/Suprime Cam and a notched  $HK'$  filter with a central wavelength of  $1.8\text{ }\mu\text{m}$  from the University of Hawaii 2.2-m telescope/QUIRC – see Fig. 1 for filter transmissions combined with the CCD responses). The resolution of the X-ray and optical/near-IR observations are similar (around  $1\text{ arcsec}$ ) with nearly all true counterparts expected to lie within the  $2\text{ arcsec}$  ( $5\sigma$ ) radii of the X-ray source for sources within  $6.5\text{ arcmin}$  of the approximate X-ray image centre and  $3.6\text{ arcsec}$  radii ( $4\sigma$ ) for sources beyond this radius. The  $1\sigma$  limiting magnitudes are (AB magnitudes):  $B$  (29.0),  $V$  (28.5),  $R$  (29.2),  $I$  (27.6),  $z'$  (27.0) and  $HK'$  (23.3, Vega). As no

errors are provided with these measurements, photometric errors of  $0.05$  in the *Subaru* bands and  $0.15$  in  $HK'$  have been assumed. The catalogue is reduced to 161 sources after removal of those flagged as saturated (seven sources) or contaminated (one source). In addition,  $HK'$  detections were only used if they were brighter than magnitude  $20.0$  for data quality reasons. This means there are 105 sources with detections in six bands, 52 with five bands and four with four bands. Dropout treatment was applied to the two shortest wavebands –  $B$  and  $V$ .

Since this sample is X-ray-selected, we can expect a large proportion to be AGN, or AGN-dominated. Hence, although this sample does not have an exactly analogous set of filters to the INT WAS, it is another excellent test-bed for the AGN template fitting. It also has a larger sample of high-redshift galaxies which test the IMPZ application to galaxies across a broader redshift range. In place of the *stellar* definition in the application to the INT WAS survey, AGN template fitting was carried out in addition to the usual galaxy template fitting for objects that are flagged as being optically compact ( $C - 23$  sources), having broad-line features ( $B - 6$  sources), or having both ( $BC - 23$  sources). Hence 52 of 161 sources have AGN template fitting applied. Barger et al. (2002) used HYPERZ on this sample but only managed to get about one-third of the photometric redshifts of their broad-line sources within 25 per cent of the spectroscopic values.

Galactic extinction in CDFN is again low, with extinction  $E(B - V) = 0.012$ , which using the extinction–wavelength relation in Cardelli et al. (1989) equates to  $A_V \approx 0.037$  for  $R_V = 3.1$ . We can also use the same relation to find the extinction in each of the bandpasses and correct for it (see Table 3).

#### 4.3 Comparison to HYPERZ

In order to see how the popular template-fitting photometric code HYPERZ compares, it has been run on the same two catalogues, using as similar parameters as possible. The set of six galaxy templates was used, with RR1UVrise and RR2UVrise used for AGN fitting. One difference between the HYPERZ and IMPZ codes is that the  $A_V$  prior that increases the  $\chi^2_{\text{red}}$  for increasing  $|A_V|$  is not implementable in HYPERZ. Similarly the prior that makes AGN fits more preferable than galaxy fits for *stellar* sources is not available in HYPERZ (these priors were set out in Section 2.2). HYPERZ was run with the same redshift range, but a redshift step in  $(1 + z)$ , rather than in  $\log_{10}(1 + z)$ . The same absolute magnitude limits as in IMPZ were used (but the equivalent limits in  $g'$  band for INT WAS) for the galaxy and AGN fits, with an  $A_V$  range of  $0.0$  to  $1.0$  for galaxy fits and  $A_V = 0$  for AGN and elliptical fits. Two reddening laws were tried, Calzetti et al. (2000) and Seaton (1979) (fitted by Fitzpatrick 1989 for the Milky Way), with galaxies getting the best results with the Seaton (1979) law. Allowing  $A_V$  to vary for AGN fits gave very similar results to setting  $A_V = 0$ , with the best results shown in Figs 4 and 6. The results of this comparison to HYPERZ are in Section 5.4.

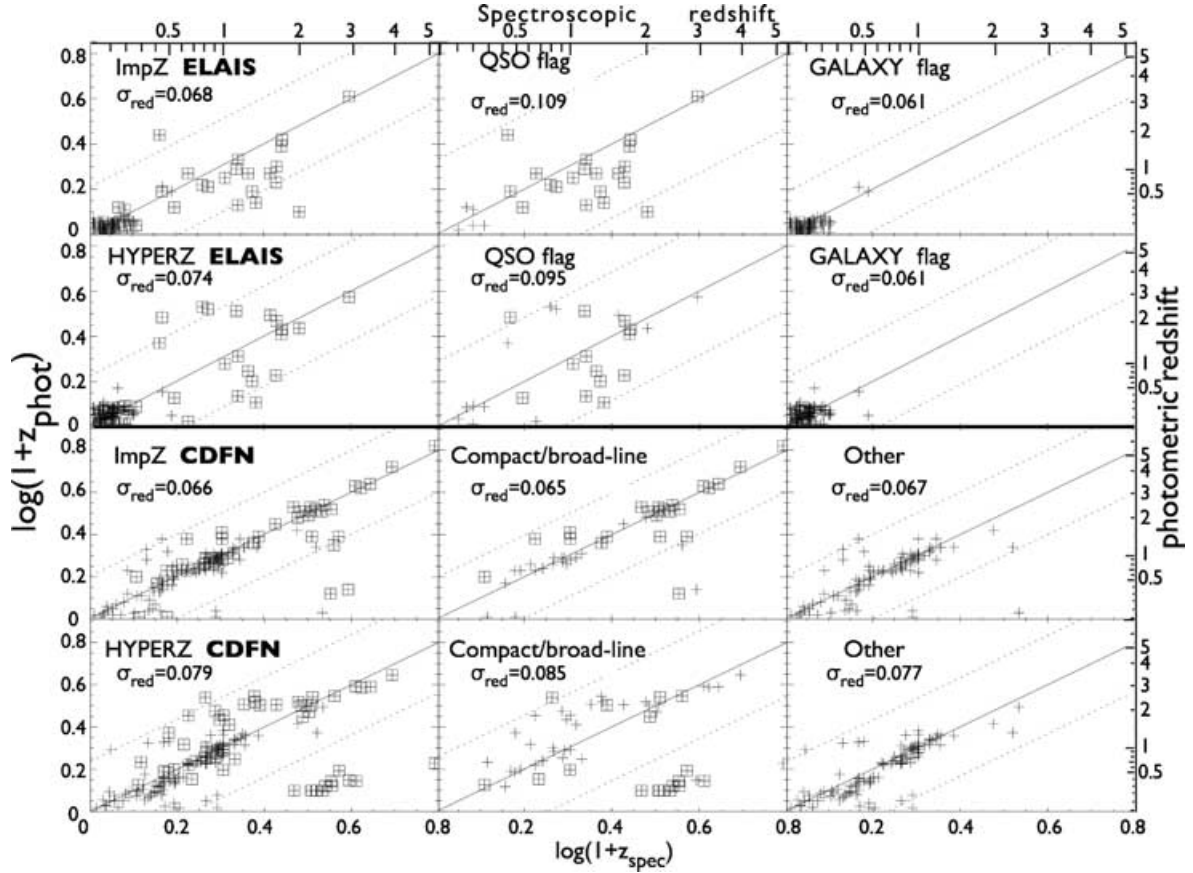
### 5 SPECTROSCOPIC REDSHIFT RESULTS

We measure the reliability and accuracy of the photometric redshifts via the fractional error  $\Delta z/(1 + z)$  for each source, examining the mean error  $\overline{\Delta z}/(1 + z)$ , the rms scatter  $\sigma_z$  and the rate of ‘catastrophic’ outliers  $\eta$ , defined as the fraction of the full sample that has  $|\Delta z/(1 + z)| > 0.2$ . The values of  $\Delta z/(1 + z)$ ,  $\overline{\Delta z}/(1 + z)$  and  $\sigma_z$  are calculated as follows:

$$\frac{\Delta z}{(1 + z)} = \left( \frac{z_{\text{phot}} - z_{\text{spec}}}{1 + z_{\text{spec}}} \right), \quad (3)$$

**Table 3.** ELAIS N1 and N2 Galactic extinction corrections for the INT WAS filters and CDFN Galactic extinction corrections for the Barger et al. (2002) filter set, derived using Cardelli et al. (1989) with  $R_V = 3.1$ .

(a) INT WAS filters						
Filter	$U$	$g'$	$r'$	$i'$	$Z$	
$A_\lambda$	0.034	0.025	0.019	0.013	0.010	
(b) Barger et al. (2002) filters						
Filter	$B$	$V$	$R$	$I$	$z'$	$HK'$
$A_\lambda$	0.049	0.037	0.031	0.022	0.018	0.005



**Figure 4.** Spectroscopic validation results for the INT WAS ELAIS and CDFN spectroscopic samples. The rms values shown are the outlier-clipped rms values,  $\sigma_{\text{red}}$ . The top two rows show INT WAS results (first row, IMPZ results; second row, HYPERZ results), set out as follows. The first plot is all the data (crosses) with those flagged as ‘QSO’ surrounded with a square. The dotted lines are the  $3\sigma_{\text{red}}$  limits. The second plot is only ‘QSO’ objects (crosses) with those fitted as AGN surrounded with a square. Dotted lines are again  $3\sigma_{\text{red}}$ . The third plot shows ‘Galaxy’ objects (crosses) with those fitted as AGN surrounded by a square. Dotted lines are  $3\sigma_{\text{red}}$  limits. The third and fourth rows show results (third row, IMPZ results; fourth row, HYPERZ results) from the CDFN sample with a similar format to the first row. The first plot is all the data (crosses) with those flagged as ‘B’, ‘C’ or ‘BC’ surrounded with a square. The dotted lines are the  $3\sigma_{\text{red}}$  limits. The second plot is only ‘B’, ‘C’ or ‘BC’ objects (crosses) with those fitted as AGN surrounded with a square. Dotted lines are again  $3\sigma_{\text{red}}$ . The third plot shows the remaining objects (crosses) with dotted lines the  $3\sigma_{\text{red}}$  limits.

$$\frac{\overline{\Delta z}}{(1+z)} = \sum \left( \frac{z_{\text{phot}} - z_{\text{spec}}}{1 + z_{\text{spec}}} \right) / N, \quad (4)$$

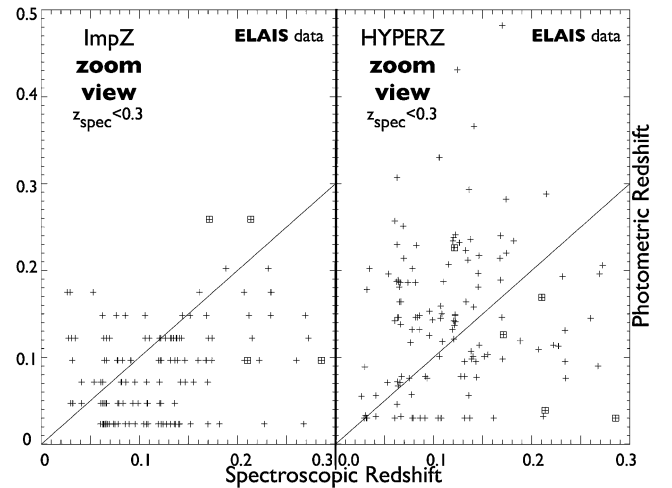
$$\sigma_z^2 = \sum \left( \frac{z_{\text{phot}} - z_{\text{spec}}}{1 + z_{\text{spec}}} \right)^2 / N, \quad (5)$$

with  $N$  being the number of sources with both spectroscopic redshifts and photometric redshifts.

For both spectroscopic samples, the outlier-clipped rms,  $\sigma_{\text{red}}$ , calculated from sources with  $|\Delta z|/(1+z) < 0.2$  was around 0.07 and the outlier fraction was 4.9 per cent for the INT WAS ELAIS and 12.4 per cent for the CDFN sample.

### 5.1 Results of ELAIS spectroscopic study

The code parameters and templates that gave the best results are described in Section 5.5. Fig. 4 shows the results of running IMPZ on the INT WAS ELAIS spectroscopic redshift sample with this ‘optimum’ set of parameters and templates, with Fig. 5 showing in close-up the behaviour at low redshift. As well as plotting  $\log_{10}(1+z_{\text{spec}})$  versus  $\log_{10}(1+z_{\text{phot}})$  for the whole sample, the results for just those with a spectroscopic ‘QSO’ flag are shown, as is the



**Figure 5.** Zoomed view of spectroscopic validation results for the INT WAS ELAIS spectroscopic sample. The first plot shows all IMPZ data for  $z_{\text{spec}} < 0.3$  (crosses) with those flagged as ‘QSO’ surrounded with a square. The second plot is the same plot for the HYPERZ results. For both plots, the scatter appears large, since the photometric redshift method is accurate to perhaps 0.05 in  $(1+z)$ , which is of the order of  $(1+z)$  for low redshifts.

‘Galaxy’ subsample. This better illustrates the varying success of the code on these different groups of sources, and the template types that were best fitting. In particular, it can be seen that 20 of the 25 ‘QSO’ objects were best-fitted by AGN templates. Four of the five that were not were defined as *non-stellar* following the definition set out in Section 2.2 (so AGN fitting was not applied to them) and all five are at low redshift ( $z < 0.3$ ).

IMPZ found solutions for all but one of the 162 sources. The failed source is discussed in Section 5.1.5.

For the sample as a whole, the total rms scatter,  $\sigma_{\text{tot}}$ , was 0.12, with  $\overline{\Delta z}/(1+z) = -0.02$ . The outlier-clipped rms,  $\sigma_{\text{red}}$ , was 0.068 and the rate of ‘catastrophic’ outliers,  $\eta$ , was 4.9 per cent.

The code was successful in identifying and fitting AGN templates to nearly all of the ‘QSO’ objects, and was quite successful at returning an accurate  $z_{\text{phot}}$ . The total rms scatter,  $\sigma_{\text{tot}}$ , was 0.29 for this subsample, with  $\overline{\Delta z}/(1+z) = -0.11$ . The outlier-clipped rms,  $\sigma_{\text{red}}$ , was 0.11 and the rate of ‘catastrophic’ outliers,  $\eta$ , was 32 per cent, higher than for the spectroscopic sample as a whole. It can be seen from Fig. 4, however, that in fact the agreement with  $z_{\text{spec}}$  is reasonable (all but one are within the  $3\sigma_{\text{red}}$  limits), especially considering the infancy of applying photometric redshifts to quasars. Nine of the 25 ‘QSO’ sources were fitted within 25 per cent of their  $z_{\text{spec}}$  values.

The code was even better on the ‘Galaxy’ sources, with  $\sigma_{\text{tot}} = 0.061$ ,  $\overline{\Delta z}/(1+z) = -0.029$  and  $\sigma_{\text{red}} = 0.061$ . There were no ‘catastrophic’ outliers. Although these ‘Galaxy’ sources were all relatively low redshift ( $z < 0.2$ ), the success is encouraging.

Statistical results for both the ‘best-case’ setup and also for other IMPZ setups can be found in Table 4.

### 5.1.1 Stellar definition

It is important to restrict the AGN templates to which the sources are fitted. If the AGN templates are fitted to all sources, instead of only fitting to those defined as *stellar*, then the overall accuracy of the IMPZ code on the INT WAS spectroscopic sample drops to  $\sigma_{\text{tot}} = 0.16$  and  $\eta$  rises to 6.8 per cent. This is because the inclusion of AGN templates introduces more degeneracies into the colour-redshift space and we therefore wish to use further information, in this case the class flags, to break some of these degeneracies.

Similarly, although the *stellar* sources are more likely to be quasars, if we restrict the fitting of galaxy templates to *non-stellar* objects only, then the accuracy of IMPZ results for *stellar* objects is reduced because not all of them will be quasars. If galaxy templates are only fitted to *non-stellar* sources, then  $\sigma_{\text{tot}} = 0.16$ ,  $\overline{\Delta z}/(1+z) = 0.01$ ,  $\sigma_{\text{red}} = 0.07$  and  $\eta = 7.5$  per cent. Instead, galaxy templates are fitted to *stellar* objects, but AGN template fits are made preferable through the prior set out in Section 2.2. The alteration of this prior is discussed in Section 5.1.4.

Bearing in mind that the code is to be applied to the INT WAS ELAIS survey as a whole where there will be proportionally less quasars and more galaxies, we wish to make the criteria for a *stellar* definition and ensuing AGN fitting as tight as possible in order to minimize galaxy contamination (and improve efficiency). The best combination of class flags for identifying quasars without undue contamination from galaxies was found to be *stellar* if the class flag is  $-1$  in  $g'$ ,  $r'$ , or  $i'$  band. This defines 21/25 ‘QSO’ sources and 3/138 ‘Galaxy’ sources as *stellar*.

### 5.1.2 AGN fitting

The choice of which AGN templates to include in the template set, and to a lesser extent how to treat them in the UV, had a large

effect on the success of the code in obtaining accurate redshifts for *stellar* objects, and also in whether the ‘QSO’ sources were correctly identified as such and best-fitted by an AGN rather than a galaxy template. At high enough redshifts for the different UV treatments to enter the INT WAS bandpasses ( $z \gtrsim 2.5$ ) there are relatively few sources available to test the different behaviours – the CDFN sample is more useful (see Section 5.2.1). For the highest-redshift INT WAS source (‘QSO’ at  $z = 2.94$ ) it was found however that the AGN templates were not the best fitting when the UVHST behaviour was used, with instead a low-redshift Sbc template being fitted.

The UVdrop behaviour was found to be reasonable, with the high-redshift source being fitted as an AGN at  $z = 3.4$  whilst the UVrise gave the redshift as  $z = 3.1$ . The UVrise behaviour was therefore chosen to be the more successful, although the main reason for its choice was determined by the CDFN results (Section 5.2), which had more high-redshift sources.

The SDSS AGN template was found to best fit a reasonable number of *stellar* sources but unfortunately the resulting photometric redshifts were inaccurate, with a larger number of catastrophic outliers. For example, using the SDSS template with the UVrise behaviour best fitted 19/25 ‘QSO’ sources, but with an increased  $\sigma_{\text{tot}} = 0.30$  and an outlier fraction of more than half.

It was found that the best combination of AGN templates was to use the RR1 and RR2 templates with the UVrise behaviour. Using just one or the other tended to increase the rms scatter of the ‘QSO’ objects from  $\sigma_{\text{tot}} = 0.28$  to roughly 0.40 and to increase the number of outliers.

It is noted that if no AGN templates are used – only the six galaxy templates – then the outlier fraction for the ‘QSO’ sources increases to  $\eta = 80$  per cent with  $\sigma_{\text{tot}} = 0.48$ . Indeed, no source with  $z_{\text{spec}} > 0.6$  is fitted within  $3\sigma_{\text{tot}}$  of  $z_{\text{spec}}$ .

### 5.1.3 High-redshift sources

Since the main power of photometric redshifts is in their application to higher redshifts, it is informative to examine the higher-redshift sources as a separate sample. Taking our redshift cut as  $z_{\text{spec}} > 0.3$  (sources below this redshift are plotted in close-up in Fig. 5) means that there are 22 sources in the high-redshift sample, comprising two ‘Galaxy’ and 20 ‘QSO’ sources. Hence the statistics for this sample are similar to that for the QSO sources alone, with  $\sigma_{\text{tot}} = 0.34$ ,  $\overline{\Delta z}/(1+z) = 0.12$  and  $\sigma_{\text{red}} = 0.11$  with an outlier rate of 36 per cent. As would be expected, then, the accuracy for higher-redshift sources is less than for lower-redshift sources, but is still good. A better grasp of high-redshift performance is provided by the CDFN sample, which has more objects at  $z_{\text{spec}} > 0.3$  – see Section 5.2.2.

### 5.1.4 Relaxation of parameters

**5.1.4.1  $A_v$ .** If  $A_v$  fitting is completely turned off, so that only  $A_v = 0$  solutions are allowed, then the majority of the increase in rms comes from increased scatter in the ‘Galaxy’ subsample. Hence the inclusion of  $A_v$  as a parameter improves the accuracy of the redshifts, whilst presumably giving some information about the extinction of each source. Several different  $A_v$  limits were tried for the galaxy templates. For example, since the ELAIS fields were originally IR-selected, we might expect some sources to have high extinction so a range of  $A_v = 0-3$  was used, with  $\beta = 3$  instead of 2 in the  $A_v$  prior that minimizes  $\chi^2_{\text{red}} + \beta A_v^2$ . This did not alter the results. However,

**Table 4.** Statistical results for application of the IMPZ code to the INT WAS ELAIS catalogue, for different IMPZ setups.

All sources				'QSO' sources				'Galaxy' sources			
$\overline{\Delta z}/(1+z)$	$\sigma_{\text{tot}}$	$\sigma_{\text{red}}$	$\eta$ (per cent)	$\overline{\Delta z}/(1+z)$	$\sigma_{\text{tot}}$	$\sigma_{\text{red}}$	$\eta$ (per cent)	$\overline{\Delta z}/(1+z)$	$\sigma_{\text{tot}}$	$\sigma_{\text{red}}$	$\eta$ (per cent)
'Best-case' IMPZ setup, as in Section 5.1											
−0.02	0.12	0.068	4.9	−0.11	0.29	0.11	32.0	−0.029	0.061	0.061	0.0
AGN fitted to all sources, as in Section 5.1.1											
−0.01	0.16	0.07	6.8	−0.05	0.36	0.11	40.0	−0.002	0.07	0.06	0.7
Galaxies only fitted to <i>non-stellar</i> sources, as in Section 5.1.1											
0.01	0.16	0.07	7.5	−0.08	0.32	0.11	36.0	0.03	0.11	0.06	2.2
SDSS template with UVrise as the only AGN template, as in Section 5.1.2											
−0.04	0.13	0.07	8.0	−0.21	0.30	0.11	52.0	−0.004	0.06	0.06	0.0
RR1 template with UVrise as the only AGN template, as in Section 5.1.2											
−0.01	0.16	0.06	6.2	−0.07	0.39	0.10	40.0	−0.004	0.06	0.06	0.0
RR2 template with UVrise as the only AGN template, as in Section 5.1.2											
−0.03	0.15	0.07	8.0	−0.17	0.36	0.12	52.0	−0.004	0.06	0.06	0.0
UVdrop behaviour in place of UVrise, as in Section 5.1.2											
−0.02	0.12	0.07	4.9	−0.11	0.29	0.12	32.0	−0.004	0.06	0.06	0.0
The four CWW templates in place of the E, Sab, Sbc, Scd and Sdm templates, as in Section 6.2											
−0.02	0.13	0.07	5.6	−0.12	0.28	0.12	32.0	−0.01	0.07	0.07	0.7
Sab, R1UVrise and RR2UVrise as the only templates, as in Section 6.2											
−0.05	0.14	0.07	8.6	−0.19	0.30	0.12	48.0	−0.03	0.07	0.06	1.5
No AGN templates, as in Section 5.1.2											
−0.07	0.20	0.06	12.3	−0.42	0.48	0.11	80.0	−0.004	0.06	0.06	0.0
$A_V$ free for AGN templates											
−0.03	0.13	0.07	6.8	−0.17	0.30	0.11	44.0	−0.004	0.06	0.06	0.0
$A_V = 0$ to 3, with $\beta = 3$ in $A_V$ prior (see Section 2.2)											
−0.02	0.12	0.07	4.9	−0.11	0.28	0.11	32.0	−0.01	0.06	0.06	0.0
$A_V = -0.4$ to 1.0											
−0.02	0.13	0.07	6.2	−0.11	0.29	0.12	36.0	−0.004	0.06	0.06	0.7
$A_V$ set to zero for all templates, as in Section 5.1.3											
0.01	0.13	0.08	5.6	−0.11	0.29	0.10	32.0	0.02	0.08	0.07	0.7
AGN prior removed, as in Section 5.1.3											
−0.01	0.15	0.07	7.4	−0.17	0.34	0.12	44.0	0.02	0.07	0.06	0.7
AGN prior set to 2 instead of 4, as in Section 5.1.3											
−0.01	0.14	0.07	6.2	−0.14	0.32	0.11	36.0	0.02	0.07	0.06	0.7
AGN prior set to 25 instead of 4, as in Section 5.1.3											
0.003	0.14	0.07	6.2	−0.08	0.32	0.11	36.0	0.02	0.07	0.06	0.7
Without $U$ band information, as in Section 6.2											
0.01	0.43	0.06	13.0	−0.11	0.43	0.11	60.0	0.03	0.43	0.06	4.4
Without $Z$ band information, as in Section 6.2											
−0.03	0.13	0.07	7.4	−0.13	0.30	0.10	48.0	−0.01	0.06	0.06	0.0
Brighter AGN limits of $-29 < M_B < -17.5$ , as in Section 6.3											
0.01	0.25	0.07	6.2	0.10	0.64	0.12	40.0	−0.004	0.06	0.06	0.0
Reducing faint AGN limits to $-27.0 < M_B < -15$ , as in Section 5.1.3											
−0.02	0.12	0.07	4.9	−0.11	0.28	0.11	32.0	−0.01	0.06	0.06	0.0
Brighter galaxy limits to $-23.5 < M_B < -13.5$ , as in Section 5.1.3											
−0.02	0.13	0.07	6.2	−0.11	0.28	0.11	32.0	−0.003	0.06	0.06	1.5
Brighter galaxy limits to $-25.5 < M_B < -13.5$ , as in Section 5.1.3											
0.01	0.33	0.06	7.4	−0.10	0.29	0.12	36.0	0.03	0.34	0.05	2.2
Reducing faint galaxy limits to $-23 < M_B < -17.5$ , as in Section 5.1.3											
−0.02	0.12	0.06	4.9	−0.11	0.28	0.11	32.0	0.001	0.06	0.06	0.0

extending  $A_v$  freedom to AGN templates increases the outlier rate for ‘QSO’s to  $\eta = 40$  per cent.

**5.1.4.2 AGN prior.** If the prior that a galaxy fit to a *stellar* source must have  $(\chi^2_{\text{red}})_{\text{galaxy}} < [(\chi^2_{\text{red}})^{\text{AGN}} - 4]$  is removed, then only 17 of 25 ‘QSO’ sources are fitted as AGN and the highest-redshift source (‘QSO’ at  $z = 2.94$ ) is instead fitted as a low-redshift Sbc galaxy, with an overall increase in the rms and outlier fraction. If the prior is reduced in strength, to  $(\chi^2_{\text{red}})_{\text{galaxy}} < [(\chi^2_{\text{red}})^{\text{AGN}} - 2]$ , then, although the overall statistics are almost as good, the highest-redshift source is again fitted as a low-redshift Sbc galaxy.

**5.1.4.3 Absolute magnitude limits.** The choice of the absolute  $B$ -band magnitude,  $M_B$ , limits of the photometric solutions is important for removing physically unlikely low-redshift low-luminosity and high-redshift high-luminosity solutions, which greatly increase the number of catastrophic outliers. For example, allowing brighter AGN limits of  $-29 < M_B < -17$  instead of  $-27 < M_B < -17$  resulted in a number of high photometric redshift ( $z > 5$ ) fits to low-redshift sources. Similarly, increasing the galaxy limits to  $-25.5 < M_B < -13$  replaced a number of correct faint low-redshift fits with incorrect brighter high-redshift fits. Incorrect low-redshift solutions for high-redshift sources were found to increase if the opposite action was taken and brighter galaxy and AGN fits were not allowed. The use of an upper  $M_B$  envelope that increased with redshift was found to be more successful than using a fixed value for galaxies, as might be expected from the known strong evolution in galaxy luminosities. The selected redshift dependence used here was  $[-22.5 - 2 \log_{10}(1+z)] < M_B < -13.5$ , though  $[-22.5 - z/3] < M_B < -13.5$  worked equally well.

The final absolute magnitude limits were chosen in order to give the best agreement with the spectroscopic data whilst taking into consideration that the INT WAS ELAIS survey as a whole will have a greater diversity of sources.

### 5.1.5 Outliers

IMPZ fails to find a solution for one INT WAS ELAIS source, a ‘Galaxy’ at redshift 0.1454 with a *non-stellar* flag. This source is not detected in  $g'$ , but is relatively bright in other bands (19.42, 18.05, 17.18, 16.75 in  $U$ ,  $r'$ ,  $i'$  and  $Z$ ). It was flagged as having multiple counterparts, so this might imply that the photometry is incorrect. HYPERZ manages to find a solution at  $z_{\text{hyp}} = 0.19$  as an Sab with  $A_v = 0.4$ .

The remaining most conspicuous outliers are a group of three sources with high spectroscopic redshifts and lower photometric redshift solutions, falling below the  $3\sigma_{\text{red}}$  boundaries. It is noted that two of these sources are at redshifts where the 4000-Å break has left the longest INT WAS waveband (at  $z \approx 1.4$ ) but the 912-Å Lyman limit has still not entered the shortest waveband ( $z \approx 2.5$ ). It is harder to derive redshifts successfully for sources in this redshift range since the primary SED features on which photometric redshift fitting relies are not available. Table 2 details the redshifts where common spectral features enter and leave the bandpasses used in the INT WAS and CDFN catalogues. For five other sources that fall into this redshift range, IMPZ is still successful in deriving the correct redshift, implying a success rate in this redshift region of 71 per cent.

The other outlier (a ‘QSO’) lies above the  $3\sigma_{\text{red}}$  boundaries, with a low  $z_{\text{spec}}$  and a high  $z_{\text{phot}}$  solution as an AGN. A solution near the correct redshift is found if the SDSS AGN template is used;

however, including this template degrades the performance of the code for the rest of the sample.

## 5.2 Results of CDFN spectroscopic study

Fig. 4 shows the results of running IMPZ on the CDFN spectroscopic redshift sample with the same ‘optimum’ set of parameters and templates as plotted for the INT WAS sample. As expected, since there is less available information on which to base the *stellar* and *non-stellar* definition and no actual ‘QSO’ or ‘Galaxy’ flag, the resulting analysis can be less clearly separated into results for galaxies and quasars.

IMPZ found solutions for all 161 sources. The total rms scatter,  $\sigma_{\text{tot}}$ , was 0.17, with  $\overline{\Delta z}/(1+z) = -0.01$ . The outlier-clipped rms,  $\sigma_{\text{red}}$ , was 0.07 and the rate of ‘catastrophic’ outliers,  $\eta$ , was 12.4 per cent. From Fig. 4 it can be seen that the code really only went badly wrong for three sources.

The code was successful in identifying and fitting AGN templates to 27/52 of the broad-line or compact objects, and overall was successful at returning an accurate photometric redshift. The total rms scatter,  $\sigma_{\text{tot}}$ , was 0.19 for this subsample, with  $\overline{\Delta z}/(1+z) = -0.01$ . The outlier-clipped rms,  $\sigma_{\text{red}}$ , was 0.07 and the rate of ‘catastrophic’ outliers,  $\eta$ , was 17.3 per cent.

The code is also successful for the remaining sources (not B and/or C sources), with  $\sigma_{\text{tot}} = 0.16$ ,  $\overline{\Delta z}/(1+z) = -0.01$  and  $\sigma_{\text{red}} = 0.07$ . The rate of ‘catastrophic’ outliers,  $\eta$ , was 10.1 per cent.

Statistical results for both the ‘best-case’ setup and also for other IMPZ setups can be found in Table 5.

### 5.2.1 AGN fitting

Nineteen of 29 (66 per cent) broad-line (B or BC) source redshifts were within 25 per cent of their spectroscopic values. As a comparison, Barger et al. (2002) got one-third of the sample within this tolerance.

If we use just BC (broad-line and compact) or C (compact) as the *stellar* definition (in place of B, C, BC), then the broad-line source numbered 174 in Barger et al. (2002) (with the highest spectroscopic redshift,  $z = 5.19$ ) is incorrectly fitted as a low-redshift galaxy. The reason why the B flag should perhaps not be used in the *stellar* definition is that this is a spectroscopically rather than photometrically derived property, and so would not be available to a photometric survey in general.

As with the INT WAS ELAIS spectroscopic sample, the most successful AGN templates were found to be the RR1 and RR2 templates and, because of the greater number of high-redshift sources in the CDFN sample, the best UV treatment could be investigated more clearly. It was found that using the UVdrop or UVHST behaviours was less successful than UVrise. The UVHST behaviour increased the population of low-redshift sources incorrectly fitted as high-redshift, although the overall success was good, with the main result being a slight increase in the outlier fraction for the *stellar* sources. The UVdrop treatment had a similar effect, making the UVrise behaviour the most effective.

It is noted that if no AGN templates are included in the template set, then the outlier rate for *stellar* sources increases to 58 per cent, with a large population of high-redshift *stellar* sources incorrectly placed at lower redshifts. If AGN templates are fitted along with galaxy templates to *non-stellar* sources, then the outlier rate for *non-stellar* sources increases to 28 per cent and  $\sigma_{\text{tot}}$  rises to 0.88. If, conversely, galaxy templates are only fitted to *non-stellar* sources, then the rms scatter of the *stellar* sources becomes much larger, with  $\sigma_{\text{tot}} = 1.42$ .

**Table 5.** Statistical results for application of the IMPZ code to the CDFN catalogue, for different IMPZ setups.

All sources				Broad-line and/or compact sources				Other sources			
$\overline{\Delta z}/(1+z)$	$\sigma_{\text{tot}}$	$\sigma_{\text{red}}$	$\eta$ (per cent)	$\overline{\Delta z}/(1+z)$	$\sigma_{\text{tot}}$	$\sigma_{\text{red}}$	$\eta$ (per cent)	$\overline{\Delta z}/(1+z)$	$\sigma_{\text{tot}}$	$\sigma_{\text{red}}$	$\eta$ (per cent)
'Best-case' IMPZ setup, as in Section 5.2											
−0.01	0.17	0.07	12.4	0.01	0.19	0.07	17.3	−0.01	0.16	0.07	10.1
AGN fitted to all sources, as in Section 5.2.1											
0.17	0.70	0.07	18.6	0.01	0.19	0.07	17.3	0.24	0.84	0.07	17.3
Galaxies only fitted to <i>non-stellar</i> sources, as in Section 5.2.1											
0.25	0.80	0.07	24.8	0.86	1.42	0.07	56.3	−0.01	0.16	0.07	11.0
SDSS template with UVrise as the only AGN template, as in Section 5.2.1											
−0.07	0.23	0.07	22.4	−0.19	0.34	0.08	48.1	−0.01	0.16	0.07	10.1
RR1 template with UVrise as the only AGN template, as in Section 5.2.1											
−0.01	0.18	0.07	13.7	0.004	0.21	0.07	21.2	−0.01	0.16	0.07	10.1
RR2 template with UVrise as the only AGN template, as in Section 5.2.1											
−0.01	0.17	0.07	12.4	0.01	0.19	0.08	17.3	−0.01	0.16	0.07	10.1
UVdrop behaviour in place of UVrise, as in Section 5.2.1											
0.06	0.52	0.07	14.3	0.21	0.88	0.08	23.1	−0.01	0.16	0.07	10.1
The four CWW templates replacing E, Sab, Sbc, Scd and Sdm templates, as in Section 6.2											
0.01	0.36	0.07	14.3	−0.05	0.20	0.08	21.1	0.04	0.42	0.06	11.0
Sab, R1UVrise and RR2UVrise as the only templates, as in Section 6.2											
0.08	0.57	0.07	36.6	0.39	0.96	0.06	59.6	−0.07	0.21	0.08	25.7
No AGN templates, as in Section 5.2.1											
−0.07	0.31	0.07	21.1	−0.18	0.49	0.08	44.2	−0.01	0.16	0.07	10.1
$A_V$ free for AGN templates, as in Section 5.2.2											
0.08	0.50	0.06	23.0	0.25	0.86	0.05	50.0	−0.01	0.16	0.07	10.1
$A_V = 0$ to 3, with $\beta = 3$ in $A_V$ prior (see Section 2.2)											
−0.02	0.18	0.08	13.7	0.01	0.20	0.08	17.3	−0.03	0.16	0.07	11.9
$A_V = -0.4$ to 1.0											
0.03	0.37	0.07	13.7	−0.01	0.20	0.07	17.3	0.05	0.43	0.07	11.9
$A_V$ set to zero for all templates, as in Section 5.2.2											
0.08	0.43	0.08	17.4	0.22	0.72	0.09	26.9	0.01	0.16	0.08	12.8
AGN prior removed, as in Section 5.2.2											
−0.03	0.19	0.07	13.7	−0.08	0.23	0.09	19.2	−0.01	0.16	0.07	11.0
AGN prior set to 2 instead of 4, as in Section 5.1.3											
−0.02	0.19	0.07	14.3	−0.03	0.23	0.08	21.1	−0.01	0.16	0.07	11.0
AGN prior set to 25 instead of 4, as in Section 5.1.3											
0.06	0.45	0.07	16.8	0.21	0.76	0.08	28.8	−0.01	0.16	0.07	11.0
Brighter AGN limits of $-29 < M_B < -17.5$ , as in Section 5.2.2											
0.13	0.68	0.07	14.9	0.43	1.17	0.07	25.0	−0.01	0.16	0.07	10.1
Reducing faint AGN limits to $-27.0 < M_B < -15$ , as in Section 5.2.2											
−0.01	0.19	0.07	13.7	−0.02	0.24	0.08	21.2	−0.01	0.16	0.07	10.1
Brighter galaxy limits to $-23.5 < M_B < -13.5$ , as in Section 5.2.2											
0.04	0.37	0.07	13.0	0.02	0.17	0.07	15.4	0.04	0.44	0.07	15.4
Brighter galaxy limits to $-25.5 < M_B < -13.5$ , as in Section 5.2.2											
0.12	0.53	0.07	18.6	0.05	0.20	0.07	19.2	0.16	0.63	0.07	18.3
Reducing faint galaxy limits to $-23 < M_B < -17.5$ , as in Section 5.2.2											
0.04	0.35	0.07	12.4	0.02	0.19	0.08	15.4	0.15	0.41	0.07	11.0

### 5.2.2 High-redshift sources

Again, we examine the higher-redshift sources as a separate sample. Taking our redshift cut as  $z_{\text{spec}} > 0.3$  (sources below this redshift

are plotted in close-up in Fig. 5) means that there are 143 sources in the high-redshift sample, comprising 92 *non-stellar* and 51 *stellar* sources. The statistics for this sample are  $\sigma_{\text{tot}} = 0.18$ ,  $\overline{\Delta z}/(1+z) = 0.01$  and  $\sigma_{\text{red}} = 0.07$  with an outlier rate of 13 per cent. The accuracy

for higher-redshift sources is therefore still good. The performance is better than for the high-redshift sample in the INT WAS ELAIS catalogue since in that case 90 per cent of the sources were ‘QSO’, for which photometric redshift fitting is less accurate.

### 5.2.3 Relaxation of parameters

**5.2.3.1  $A_v$ .** If  $A_v$  fitting is completely turned off, so that only  $A_v = 0$  solutions are allowed, then  $\eta$  increases to 17.4 per cent. Again, the inclusion of  $A_v$  as a parameter improves the accuracy of the redshifts, though analysis of a sample with known  $A_v$  would be required to quantify how the resulting  $A_v$  of the solution compares to actual  $A_v$ .

**5.2.3.2 AGN prior.** If the prior that a successful galaxy fit to a *stellar* source must have  $(\chi^2_{\text{red}})_{\text{galaxy}} < [(\chi^2_{\text{red}})^{\text{AGN}} - 4]$  is removed, then only 13 of the 52 *stellar* sources are fitted by AGN and many of those that are then fitted as galaxies are placed at much lower  $z_{\text{phot}}$ . Using a reduced prior of  $(\chi^2_{\text{red}})_{\text{galaxy}} < [(\chi^2_{\text{red}})^{\text{AGN}} - 2]$  causes the highest  $z_{\text{spec}}$  source again to be fitted as a low-redshift galaxy, and increasing the prior to  $(\chi^2_{\text{red}})_{\text{galaxy}} < [(\chi^2_{\text{red}})^{\text{AGN}} - 25]$  creates a number of false high- $z_{\text{phot}}$  AGN solutions.

**5.2.3.3 Absolute magnitude limits.** As with the INT WAS sample, allowing brighter AGN limits of  $-29 < M_B < -17.5$  instead of  $-27 < M_B < -17.5$  resulted in a number of high photometric redshift ( $z > 5$ ) fits to low-redshift sources. Similarly, increasing the luminous galaxy limits to  $-25.5 < M_B < -13.5$  replaced a number of correct faint low-redshift fits with incorrect luminous high-redshift fits. Incorrect low-redshift solutions for high-redshift sources were found to increase if the opposite action was taken and brighter galaxy and AGN fits were not allowed. Again, allowing an increase in the upper galaxy  $M_B$  envelope with redshift was found to be the best solution [i.e.  $-22.5 - 2 \log_{10}(1 + z)$ ].

### 5.2.4 Outliers

There are a group of three high spectroscopic redshift sources whose photometric solutions are significantly lower than their spectroscopic redshifts (there are a total of six sources below the  $3\sigma_{\text{red}}$  boundary, and one above, but the others are close to the boundaries). It is noted that this group of three are at redshifts where the 4000-Å break has left the Z waveband (at  $z \approx 1.5$ ) but the 912-Å Lyman limit has still not entered the shortest waveband ( $z \approx 3.3$ ). It is harder to derive redshifts successfully for sources in this redshift range since the primary SED features on which photometric redshift fitting relies are not available. For 20 other sources that fall into this redshift range, IMPZ is still successful in deriving the correct redshift, implying a success rate in this redshift region of 87 per cent. The slightly better performance than for the INT WAS ELAIS sample (71 per cent) for the corresponding ‘feature desert’ is likely to be partly due to the inclusion of *HK* information (which 4000 Å enters at  $z \approx 2.7$ ) and partly due to the poor number statistics for the INT WAS ELAIS sample (only six sources in that redshift range).

## 5.3 Statistical properties

The distribution of residuals of the IMPZ solutions in Fig. 6 is quite strongly peaked around zero, with a Gaussian-like distribution and a slightly non-Gaussian extension to larger values. The distributions show that the IMPZ code is successful to an accuracy of perhaps

0.1–0.2 in  $(1 + z)$  for nearly all sources. The distributions of IMPZ solutions (dashed line in Fig. 6) for the ‘QSO’ subsample is also Gaussian-like, but these distributions are wider and shallower. For the INT WAS ELAIS ‘QSO’s, the distribution is peaked at around  $-0.2$  rather than zero, implying that the redshift solutions tend to be slightly larger than the true values. This could be because  $A_v$  is set to zero for AGN, so a QSO reddened by some extinction is fitted as unreddened, but at a slightly higher redshift. However, this is not seen in the larger CDFN sample.

The  $\chi^2_{\text{red}}$  distributions of the IMPZ solutions show that the majority of solutions are good, with  $\chi^2_{\text{red}}$  peaking below 2. There is no large population of poor  $\chi^2_{\text{red}}$  solutions. From the IMPZ type distributions it can be seen that Sab solutions dominate, with a sizeable fraction of Scd and starbursts and very few Sdm solutions.

The  $A_v$  distributions are strongly peaked at zero with an even spread of higher values (but no clustering at the highest value of  $A_v = 1$ , which would imply that the  $A_v$  range was insufficient).

## 5.4 HYPERZ results

From Figs 4 and 6 it can be seen that, for both the INT WAS and CDFN spectroscopic samples, the IMPZ code manages to do better. The strongest cause of these differences is probably the AGN and  $A_v$  priors that are used in the IMPZ code. In particular, HYPERZ is not as successful for the broad-line and/or compact sources in the CDFN sample, with a larger population of high-redshift sources incorrectly placed at low-redshift (see Fig. 4).

Overall, HYPERZ has, for the INT WAS sample,  $\sigma_{\text{tot}} = 0.17$ ,  $\overline{\Delta z}/(1 + z) = 0.02$ ,  $\sigma_{\text{red}} = 0.073$  and  $\eta = 8.0$  per cent. For the CDFN sample the statistics were  $\sigma_{\text{tot}} = 0.26$ ,  $\overline{\Delta z}/(1 + z) = -0.04$ ,  $\sigma_{\text{red}} = 0.08$  and  $\eta = 24.2$  per cent. Only 9/29 broad-line source redshifts were within 25 per cent of their spectroscopic values.

From Fig. 6 it can be seen that the non-Gaussian tails of the redshift residuals are more pronounced, particularly for the CDFN sample, and the  $\chi^2_{\text{red}}$  of the solutions are only weakly peaked to low values.

A consideration of the ‘feature desert’ in redshift space where the 4000-Å break has left the optical wavebands, and the 912-Å Lyman limit has not yet entered, shows a marked difference from the IMPZ results for the CDFN. Indeed it is this region where HYPERZ most strongly diverges from IMPZ in its outlier rate: for the CDFN sample, IMPZ had a 87 per cent success rate with the 23 sources in this range, whereas HYPERZ fails for 10 (57 per cent success).

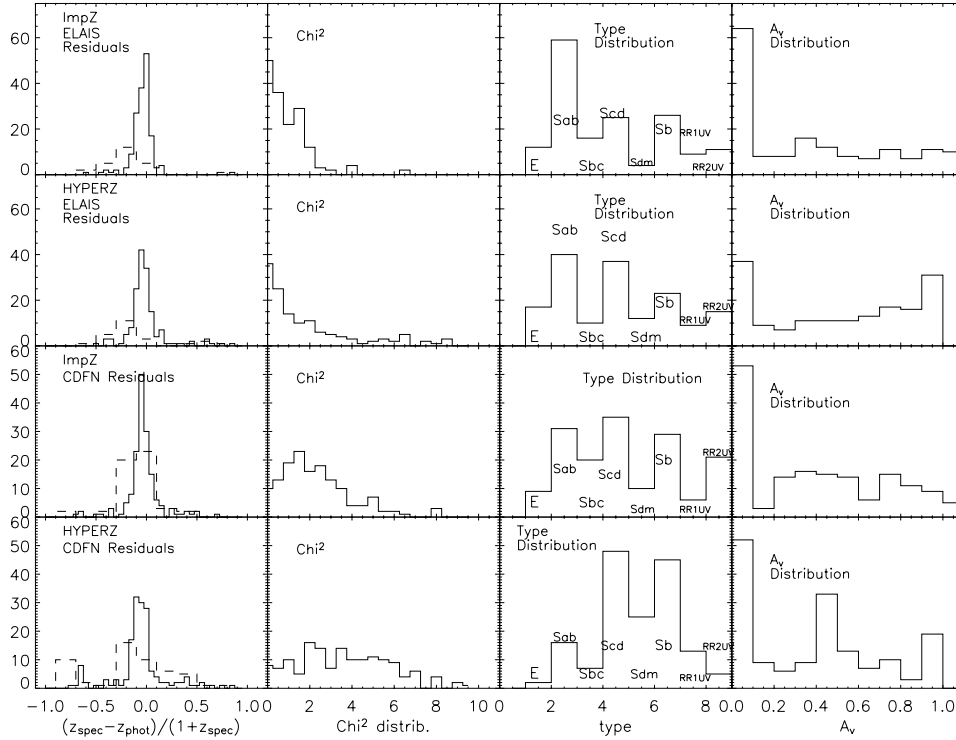
## 5.5 IMPZ summary parameters

The final best-case setup for application of IMPZ to the INT WAS ELAIS as a whole is as follows:

**Templates.** Six galaxy templates (E1, Sab, Sbc, Scd, Sdm and Sbrst) and two AGN templates (RR1UVrise and RR2UVrise), with IGM and Galactic extinction corrections.

**Fitting.** AGN templates fitted to *stellar* sources only and galaxy templates fitted to all sources, with the prior that, if the source is *stellar*, a successful galaxy fit must have  $(\chi^2_{\text{red}})_{\text{galaxy}} < [(\chi^2_{\text{red}})^{\text{AGN}} - 4]$ . Freedom in  $A_v$  is allowed for Sab, Sbc, Scd, Sdm and Sbrst templates, with a prior that makes low  $A_v$  fits preferable.

**Limits.** Limits for  $A_v$  of 0.0 to 1.0 and absolute magnitude limits of  $[-22.5 - 2 \log_{10}(1 + z)] < M_B < -13.5$  for galaxies and  $-27.0 < M_B < -17.5$  for AGN.



**Figure 6.** Spectroscopic validation results for the INT WAS ELAIS and CDFN spectroscopic samples. The top two rows show INT WAS results (first row, IMPZ results; second row, HYPERZ results), set out as follows. The first plot is the distribution of residuals between  $z_{\text{spec}}$  and  $z_{\text{phot}}$ , normalized to  $(1 + z_{\text{spec}})$ . The solid histogram is for the whole sample, the dashed histogram is that of QSO objects only. The second plot shows the reduced  $\chi^2$  distribution of the fits. The third plot is the type distribution of the fits, where 1 = E, 2 = Sab, 3 = Sbc, 4 = Scd, 5 = Sdm, 6 = Sb, 7 = RR1UVrise and 8 = RR2UVrise. The fourth plot is the  $A_v$  distribution. The third and fourth rows show results (third row, IMPZ results; fourth row, HYPERZ results) from the CDFN sample with the same format as the first two rows, except that the dashed histogram is now that of broad-line and/or compact sources only.

## 6 ERROR ANALYSIS

Errors in the derived photometric redshifts arise from several causes. The first is the inherent error in the measured flux of sources, which is expected to be increasingly important for fainter sources. The template-fitting technique adds further error due to fitting the continuum of observed galaxy SEDs with a set of standard templates that can only sample this continuum. This is known as cosmic variance.

Catastrophic errors are usually due to degeneracies in the colour–redshift–extinction space. For example, a late-type, heavily extinguished galaxy may appear similar to an early-type unreddened galaxy, but have a very different redshift. The relative magnitude of these sources of error needs to be more properly quantified.

### 6.1 Photometry

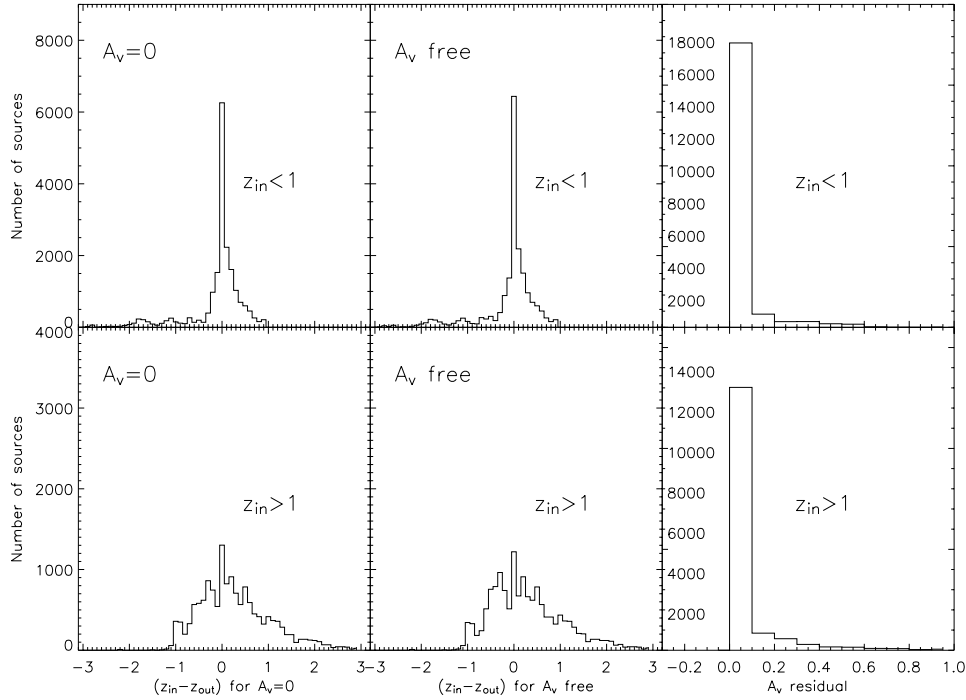
In order to remove the effect of cosmic variance from the analysis, a synthetic catalogue of fluxes ( $U$ ,  $g'$ ,  $r'$ ,  $i'$  and  $Z$ ) was generated five times, for 7000 hypothetical *sources*. These *sources* were created by redshifting the templates (equal numbers of each type) and calculating their fluxes in each band. The redshift distribution was chosen to resemble that of ‘main’ galaxies in the SDSS First Data Release, with a simple Gaussian distribution peaked at  $-20$  with a  $\sigma$  of 7 to model the absolute magnitude ( $r'$ ) distribution of the *sources* (additionally, *sources* were prevented from having extreme absolute magnitudes). The fluxes were randomly altered in a Gaussian distribution of errors of width determined by the  $1\sigma$  values of the photometric errors in the INT WAS ELAIS catalogue plus 1 per cent of the flux. This Monte Carlo simulation generates 35 000 different

*sources*. The photometric redshift code is then used to predict, for  $A_v = 0$ , the redshifts and spectral types of the *sources*. Since the synthetic catalogue is constructed from the same set of templates as those used in the IMPZ code, cosmic variance is eliminated, leaving only the effect of photometric error.

Fig. 7 shows the result of this procedure. The first plot (both rows) shows the distribution of residuals between  $z_{\text{in}}$  and  $z_{\text{out}}$ , with the first row showing results for low-redshift input sources ( $z_{\text{in}} < 1$ ) and the second row showing results for high-redshift input sources ( $z_{\text{in}} > 1$ ). The distribution of error in redshift for the low- $z_{\text{in}}$  sample is very strongly peaked at zero and is relatively symmetric about this, but the low-lying extension to larger errors is not that of a purely Gaussian distribution. These large shifts in the redshift occur when a secondary minimum in the  $\chi^2_{\text{red}}$  distribution for the *source* fluxes becomes the primary minimum when the *source* fluxes are perturbed. Hence for the low- $z_{\text{in}}$  sample, it can be seen that, for the large majority of *sources*, the redshift solution is close to the correct value, with only a low-level Gaussian spread due to photometric error.

For the high- $z_{\text{in}}$  sample the distribution is also quite symmetrical about zero, but is less strongly peaked. This is to be expected because of the poorer signal-to-noise ratio photometry of these more distant sources. For the high- $z_{\text{in}}$  sample, then, the dominant effect is the error in the photometry, causing the general broadening in the residual distribution.

The spectral types of the simulated galaxies are compared with the predicted types in Table 6. This gives the percentage of galaxies for which their spectral type in the synthetic catalogue was successfully reproduced for each template.



**Figure 7.** Results of Monte Carlo synthetic catalogue investigations with 35 000 *sources* (Section 6). The top row shows results for ‘low’ input redshift *sources* (those with  $z_{\text{in}} < 1$ ), set out as follows. The first plot is the distribution of residuals between  $z_{\text{in}}$  and  $z_{\text{out}}$ , for template fitting by IMPZ with  $A_v$  freedom turned off. The second plot is the same, but for solutions with  $A_v$  freedom allowed in IMPZ fitting, as described in Section 5.5. The third plot shows the  $A_v$  residual from fitting with  $A_v$  freedom turned on. This is simply the  $A_v$  of the fit since the input synthetic catalogue was generated without  $A_v$ . The second row displays the same information for the ‘high’ input redshift *sources* (those with  $z_{\text{in}} > 1$ ).

**Table 6.** Comparison between the input model template types and the best-fitting templates when IMPZ was run on the Monte Carlo generated model catalogues. Values are given as the percentage for each input template that was fitted by each output template. Ideally only the diagonal region (bold) would be populated.

Out	Input SED type							
	E	Sab	Sbc	Scd	Sdm	Sbrst	RR1	RR2
E	<b>88</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sab	8.0	<b>71.7</b>	18.6	0.0	2.9	0.0	0.1	0.3
Sbc	2.0	8.3	<b>71.2</b>	0.0	0.0	0.0	0.0	0.0
Scd	2.0	8.3	10.2	<b>84.8</b>	4.3	7.4	0.0	0.0
Sdm	0.0	10.0	0.0	3.0	<b>87</b>	25.0	0.0	0.0
Sbrst	0.0	1.7	0.0	12.2	5.8	<b>67.6</b>	3.9	4.7
RR1	0.0	0.0	0.0	0.0	0.0	0.0	<b>64.3</b>	30.3
RR2	0.0	0.0	0.0	0.0	0.0	0.0	31.7	<b>64.7</b>

From this analysis we conclude that, without  $A_v$  freedom, the redshifts of the simulated *sources* are well reproduced within 0.1 in  $(z_{\text{in}} - z_{\text{out}})$  for  $0 < z < 1$  and roughly 0.4 in  $(z_{\text{in}} - z_{\text{out}})$  for  $z > 1$ . Also, for low-redshift *sources*, photometric errors have a reduced effect on the results, whereas for high-redshift *sources* the precision of the photometry has the major influence on the redshift accuracy.

## 6.2 Templates

In order to investigate the effect of the choice of template set on the photometric redshifts, the four UV-extended CWW templates (E, Sbc, Scd, Im) are used in place of five galaxy templates in the IMPZ code (E1, Sab, Sbc, Scd, Sdm), retaining the starburst template

and the two AGN templates, RR1UVrise and RR2UVrise. The code is then applied to INT WAS ELAIS and CDFN spectroscopic catalogues. The resulting photometric redshifts are compared to the output when the original templates were used. The results are almost as good, with  $\sigma_{\text{tot}} = 0.13$ ,  $\sigma_{\text{red}} = 0.07$  and  $\eta = 5.6$  per cent for the INT WAS sample and  $\sigma_{\text{tot}} = 0.36$ ,  $\sigma_{\text{red}} = 0.07$  and  $\eta = 14.3$  per cent for the CDFN. The six galaxy templates were also replaced with just the Sab template, in order to represent an undersampling of SED space. It was found that, even in this rather extreme example,  $\sigma_{\text{tot}} = 0.14$ ,  $\sigma_{\text{red}} = 0.07$  and  $\eta = 8.6$  per cent for the INT WAS sample, and  $\sigma_{\text{tot}} = 0.57$ ,  $\sigma_{\text{red}} = 0.07$  and  $\eta = 36.6$  per cent for the CDFN.

It is clear that the template-fitting procedure is largely unaffected by the specific choice of templates to use, provided that the main SED features of normal galaxies (such as the Balmer break) are represented.

Table 6 shows the comparison between the input *source* template and the template that was fitted for it for the 35 000 *sources* generated in Section 6.1. It is immediately clear that the majority (75 per cent) of the input template types are recovered, and that there is almost no degeneracy between early- and late-type templates. There is also a clear demarcation between the galaxy templates and the AGN templates – very few AGN templates are fitted as galaxy templates (with some overlap with the starburst template). Recall that the reverse does not occur (galaxy template fitted by AGN) since *sources* with an input AGN template are treated as *stellar* by IMPZ. Although there are some degeneracies between the templates, this does not affect the photometric redshift accuracy greatly (as can be seen by the redshift residuals in Fig. 7). Hence the photometric redshift of a galaxy is more certain than its exact spectral type, as expected if the bulk of redshift identification is due to common features such

as the Balmer break and in agreement with the results of fitting the INT WAS ELAIS catalogue with only one galaxy template, and also from the study of Bolzonella et al. (2000).

### 6.3 Filters

What are the relative importance of different filters? In order to derive redshifts successfully for different sources, the filter set in use needs to have sufficient wavelength coverage to encompass the main template features for a broad range of redshifts. As seen in Sections 5.1.5 and 5.2.4, most outliers were sources whose spectroscopic redshifts meant that the main template features (Balmer break and Lyman limit) did not fall within filter bandpasses. It is of interest to see how results change when a filter is not included in the template fitting. To this end, IMPZ was run on the INT WAS ELAIS catalogue without *U*-band information. The statistical results of this can be found in Table 4, but the main effect was that, without *U*-band photometry, all  $z_{\text{phot}}$  values were below 1.5, although seven sources lie at  $z_{\text{spec}} > 1.5$ . This is because information on the position of the  $\text{Ly}\alpha$  line has been lost, and at these  $z_{\text{spec}}$  redshifts no other main template feature falls into the INT WAS bands.

Additionally, a synthetic catalogue of 5000 *sources* was generated as in Section 6.1 but without *U*-band photometry. For the  $z_{\text{in}} < 1$  sample, the  $(z_{\text{in}} - z_{\text{out}})$  residual distribution was strongly peaked at zero, in a similar fashion to the first plot in Fig. 7, but with a population of outliers with  $(z_{\text{in}} - z_{\text{out}}) \approx -2.6$ . This suggests that, without the use of the *U* band to decide if the Lyman break feature is present or not, the Balmer break can be mistaken to be  $\text{Ly}\alpha$  in the  $g'$  band. Hence the low- $z_{\text{in}}$  *source* is then placed at a high  $z_{\text{out}}$ . As a whole, though, the loss of *U*-band information does not adversely effect the  $z_{\text{in}} < 1$  sample. In contrast, the  $z_{\text{in}} > 1$  sample is highly dependent on it for a reasonable  $z_{\text{out}}$  value. The  $(z_{\text{in}} - z_{\text{out}})$  residual distribution is almost flat – the  $z_{\text{out}}$  accuracy is very poor. Thus the *U* band is central to good photometric redshift accuracy for  $z \gtrsim 1.5$ . This is clear from Table 2 – without *U* band, none of the template features listed is present in the remaining INT WAS filters until  $z > 2.4$  when  $\text{Ly}\alpha$  enters the  $g'$  band.

#### 6.3.1 Degeneracies in parameter space

As seen in Sections 6.1 and 6.2 there do exist degeneracies in the parameter space and this is a cause of outliers in the photometric error analysis of the simulated catalogues. Common degeneracies include confusion between the main template features (when either the redshift or the lack of bands means that most features do not fall into the filter bandpasses) and confusion between different galaxy templates. Often, however, the spectral types can be degenerate without altering  $z_{\text{phot}}$  greatly (compare Table 6 with Fig. 7).

It is also of interest to see if  $A_v$  introduces its own degeneracies. Such degeneracies will be inherent even if  $A_v$  is not explicitly fitted, since real galaxies will have different  $A_v$ . In order to quantify this, the same synthetic catalogue procedure as in Section 6.1 is carried out, but the IMPZ code is allowed to fit for free  $A_v$ . Since the input catalogue is created with  $A_v = 0$ , we can see how fitting for free  $A_v$  alters the findings.

Fig. 7 shows the result of this procedure as the second and third plots on each row (again, the first row is for the  $z_{\text{in}} < 1$  sample and the second row is for the  $z_{\text{in}} > 1$  sample). The distributions of residuals between  $z_{\text{in}}$  and  $z_{\text{out}}$  are very similar to those for the  $A_v = 0$  case, with both the low- $z_{\text{in}}$  and high- $z_{\text{in}}$  samples being strongly peaked at zero. The low- $z_{\text{in}}$  sample has a noticeable but low-level tail to higher residuals, whilst the high- $z_{\text{in}}$  sample can be well represented

by a broad Gaussian. The high- $z_{\text{in}}$  sample has a slight secondary bump at  $(z_{\text{in}} - z_{\text{out}}) \approx -0.4$ , which is likely to be due to confusion between the Balmer break and O [III] doublet, since, as can be seen in Table 2, these features enter and leave the  $i'$  and *Z* band with a separation of around 0.4 in redshift.

The  $A_v$  residual is also shown. This is simply the  $A_v$  of the fit since the input synthetic catalogue was generated without  $A_v$ . The low- $z_{\text{in}}$  sample is very strongly peaked at zero, and the high- $z_{\text{in}}$  sample is also strongly peaked at zero, with slightly more  $A_v > 0$  solutions. This means that the increasing photometric error for more distant/faint *sources* causes a few *sources* to be fitted as redder *sources* than they really are. This however does not necessarily cause an increase in the redshift error, as can be seen from the redshift residual plots for the  $A_v$ -free fitting. It is clear that the addition of  $A_v$  freedom introduces little degeneracy, and the fact that the redshift solutions tend to improve lends weight to the argument that the  $A_v$  thus determined reflects the physical  $A_v$  of the galaxy.

Running the Monte Carlo simulations with larger photometric errors tended to increase the width of the residual distributions.

### 6.4 Overall errors

The main sources of error are due to photometry, cosmic variance and degeneracies in colour–redshift–extinction space. The relative effect of each depends on the signal-to-noise ratio of the source, with photometric errors dominating those with low signal-to-noise ratio. Splitting into low ( $z < 1.5$ ) and high redshift ( $z > 1.5$ ), we can broadly say that the error for low redshifts is 0.1 or less in  $(1 + z)$ , whereas for high redshifts the error can be as large as 0.3 or more.

## 7 DISCUSSION

A comparison can be made with the photometric accuracies of other recent works: Photometric redshifts for galaxies in the Great Observatories Origins Deep Survey (GOODS; Dickinson et al. 2001) were estimated for 434 galaxies by Mobasher et al. (2004), where in total there were as many as 18 independent photometric measurements for each galaxy. Using the Bayesian method (Benítez 1999), they obtained a total rms scatter,  $\sigma_{\text{tot}}$ , of 0.11, with an outlier-clipped rms,  $\sigma_{\text{red}}$ , of 0.047 and an outlier fraction,  $\eta$ , of roughly 10 per cent. This is very similar to the success of the IMPZ code on the INT WAS ELAIS sample, which had a maximum of only five photometric bands. In Richards et al. (2001b), using statistics based on  $\Delta z = (z_{\text{phot}} - z_{\text{spec}})$ , redshifts were estimated for 2625 quasars in the SDSS filter system with an overall rms,  $\sigma_z$ , of 0.676 and  $\sigma_{\text{red}}$  of 0.1 for sources with  $|\Delta z| < 0.3$ .

The success of the IMPZ code can also be directly compared for quasars to the results of Barger et al. (2002), who got one-third of broad-line sources within 25 per cent of  $z_{\text{spec}}$ . IMPZ fitted two-thirds of the same CDFN broad-line sample within 25 per cent. More recently, Kitsionas et al. (2003) estimated photometric redshifts for a sample of X-ray-selected QSOs in the SDSS bands (up to five filters), obtaining  $z_{\text{phot}}$  within  $|\Delta z| < 0.3$  for 20 of the 30 QSOs at  $z > 0.4$ . The same statistic can be calculated for the INT WAS ELAIS ‘QSO’  $z > 0.4$  subsample, where 10 of 20 sources are within the same  $\Delta z$  tolerance, and for the broad-line and/or compact CDFN sources at  $z > 0.4$ , where 36 of 50 (72 per cent) of sources are within  $|\Delta z| < 0.3$ . The IMPZ quasar redshift results are clearly highly successful and show that reliable redshift estimates for quasars can be achieved.

It is important to note that the intrinsic variability of quasars offsets the magnitudes measured in various bands depending on the

actual epoch of observation. This is certainly an issue for the INT WAS survey, which was completed over several years. The variability scrambles the real spectrophotometric data taken over an extended period whereas the template fits assume that the photometry in each band is taken at the same time (or that there is no variability) and is therefore sampling the same underlying SED. This problem was noted in Wolf et al. (2001), who found that around half their quasar sample had completely incorrect photometric redshifts, presumably mainly due to this problem. Clearly, future surveys need to be scheduled with this effect in mind.

The SWIRE survey (in progress) combines both ground-based optical photometry and IR data (3.6, 4.5, 5.6, 8, 24, 70 and 160  $\mu\text{m}$ ). Virtually all SWIRE galaxies are expected to be detected in the IRAC 3.6- $\mu\text{m}$  band. Since little contribution from dust is expected at this wavelength, this band can be used to improve the photometric redshift estimates. Additionally, at least for low-redshift galaxies, an estimate can be made of the total stellar mass in the galaxy.

Where galaxies are detected at 4.5, 5.8 and 8  $\mu\text{m}$  an estimate can also be made of the luminosity in the cirrus emission component, after subtraction of the predicted starlight contribution. The ratio of  $L_{\text{cirr}}$  to, say,  $L_B$  should be related to the  $A_V$  value.

The crucial improvement of a survey such as SWIRE over deep fields such as HDF will be the far larger volume sampled, so that cosmic variance can be properly addressed. Although photometric redshift techniques have been applied successfully to the HDF, their true statistical power comes from application to large-area surveys such as SWIRE, which have large numbers of galaxies. Application of the IMPZ code to the SWIRE survey will be presented in a future work.

## 8 CONCLUSIONS

In this work, photometric redshifts have been studied using SED template fitting on two spectroscopic redshift catalogues (INT WAS ELAIS and CDFN). The overall accuracy of the redshift code, IMPZ, is found to be good, with  $\Delta z/(1+z) < 0.1$  for 92 per cent of galaxies (where this statistic is calculated from the 138 ‘Galaxy’ sources in the INT WAS ELAIS catalogue). The addition of  $A_V$  freedom in the fitting improves the redshift solution, suggesting that some information about the true  $A_V$  of the source is also returned. Application to sources with known  $A_V$  will help to quantify this information (Babbedge, in preparation). The template-fitting method is also extended to quasars via the inclusion of AGN templates and is found to be reasonable, with  $\Delta z/(1+z) < 0.25$  for 68 per cent of quasars (based on the 25 ‘QSO’ sources in the INT WAS ELAIS catalogue). It is noted that the inherent time variability of quasars reduces the effectiveness of the technique for surveys where the photometry is collected over an extended period.

This is the first time that a combined galaxy–quasar approach to template-fitting photometric redshift techniques has been presented, and the results and analysis presented in this work clearly show that photometric redshifts can be calculated with good success for *both* galaxies and quasars, as long as certain considerations are taken into account: It is important to limit the absolute magnitude of the solutions to prevent unphysical results; and to use available optical morphology/stellarity information in order to pre-select sources that are more likely to be quasars than galaxies (since we expect quasars to be point-like whereas most galaxies will be extended in some way). Although the application of template fitting to a single source in order to derive its redshift can never be completely relied upon, because of the inherent degeneracies and limitations of a method

reliant on perhaps four or five data points, the statistical information drawn from application to a large catalogue of objects is extremely powerful and can be used for many different investigations.

The IMPZ code will be applied to the entire recalibrated INT WAS ELAIS N1 and N2 data in order to investigate the evolution of extinction and star formation rates in a companion to this paper (Babbedge et al., in preparation).

## ACKNOWLEDGMENTS

We thank the anonymous referee for insightful comments that have served to strengthen this paper. We wish to thank Jacopo Fritz, who made the major contribution to the development of the spectrophotometric synthesis code adopted to produce the optical–NIR templates. We are also grateful Duncan Farrah for his helpful comments.

The INT WAS data is made publicly available through the Isaac Newton Groups’ Wide Field Camera Survey Programme. The Isaac Newton Telescope is operated on the island of La Palma by the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho and the Max Planck Society. The SDSS website is <http://www.sdss.org/>. We also thank A. J. Barger and co-workers for making their CDFN catalogue publicly available, and M. D. Gregg and M. Lacy for providing the optical and IR spectrum of a red quasar.

MP acknowledges support from the *Spitzer* Legacy Science Program, funded through support provided by NASA and issued by the Jet Propulsion Laboratory, California Institute of Technology. IPF was supported in part by grant PB1998-0409-C02-01 of the Spanish Ministerio de Ciencia y Tecnología. EGS acknowledges support by the Marie Curie Fellowship MCFI-2001-01809.

## REFERENCES

- Abazajian K. et al., 2003, *AJ*, 126, 2081
- Ball N. M., Loveday J., Fukugita M., Nakamura O., Okamura S., Brinkmann J., Brunner R. J., 2004, *MNRAS*, 348, 1038
- Barger A. J., Cowie L. L., Brandt W. N., Capak P., Garmire G. P., Hornschemeier A. E., Steffen A. T., Wehner E. H., 2002, *AJ*, 124, 1839
- Benítez N., 1999, in Weymann R. J., Storrie-Lombardi L. J., Sawicki M., Brunner R. J., eds, *ASP Conf. Ser. Vol. 191, Photometric Redshifts and the Detection of High Redshift Galaxies*. Astron. Soc. Pac., San Francisco, p. 31
- Benítez N., 2000, *ApJ*, 536, 571
- Berta S., Fritz J., Franceschini A., Bressan A., Pernechele C., 2003, *A&A*, 403, 119
- Berta S., Fritz J., Franceschini A., Bressan A., Lonsdale C., 2004, *A&A*, 418, 913
- Bolzonella M., Miralles J.-M., Pelló R., 2000, *A&A*, 363, 476
- Brandt W. N. et al., 2001, *AJ*, 122, 2810
- Bressan A., Granato G. L., Silva L., 1998, *A&A*, 332, 135
- Brodwin M., Lilly S., Crampton D., 1999, in Weymann R. J., Storrie-Lombardi L. J., Sawicki M., Brunner R. J., eds, *ASP Conf. Ser. Vol. 191, Photometric Redshifts and the Detection of High Redshift Galaxies*. Astron. Soc. Pac., San Francisco, p. 105
- Brotherton M. S., Tran H. D., Becker R. H., Gregg M. D., Laurent-Muehleisen S. A., White R. L., 2001, *ApJ*, 546, 775
- Bruzual A. G., Charlot S., 1993, *ApJ*, 405, 538
- Budavári T. et al., 2001, *AJ*, 122, 1163
- Calzetti D., Kinney A. L., 1992, *ApJ*, 399, L39

- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
- Carballo R., González-Serrano J. I., Benn C. R., Sánchez S. F., Vigotti M., 1999, *MNRAS*, 306, 137
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245
- Coleman G. D., Wu C.-C., Weedman D. W., 1980, *ApJS*, 43, 393 (CWW)
- Cristiani S., Vio R., 1990, *A&A*, 227, 385
- Dickinson M. et al. (GOODS Legacy Team), 2001, *BAAS*, 33, 820
- Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, *PASP*, 110, 761
- Fioc M., Rocca-Volmerange B., 1997, *A&A*, 326, 950
- Fitzpatrick E., 1989, in Allamandola L.J., Tielens A.G.G.M., eds, *Proc. IAU Symp. 135, Interstellar Dust*. Reidel, Dordrecht, p. 37
- Francis P. J., 1996, *PASP*, 13, 212
- Francis P. J., Hewett P. C., Foltz C. B., Chaffee F. H., Weymann R. J., Morris S. L., 1991, *ApJ*, 373, 465
- Giallongo E., D’Odorico S., Fontana A., Cristiani S., Egami E., Hu E., McMahon R. G., 1998, *AJ*, 115, 2169
- Gregg M. D., Lacy M., White R. L., Glikman E., Helfand D., Becker R. H., Brotherton M. S., 2002, *ApJ*, 564, 133
- Gwyn S. D. J., 1999, in Weymann R. J., Storrie-Lombardi L. J., Sawicki M., Brunner R. J., eds, *ASP Conf. Ser. Vol. 191, Photometric Redshifts and the Detection of High Redshift Galaxies*. Astron. Soc. Pac., San Francisco, p. 61
- Hatziminaoglou E., Mathez G., Pelló R., 2000, *A&A*, 359, 9
- Hogg D. W. et al., 1998, *AJ*, 115, 1418
- Jaynes E. T., 2000, *Probability Theory: The Logic of Science*. Cambridge Univ. Press, Cambridge
- Kitsionas E., Hatziminaoglou E., Georgantopoulos I., Georgakakis A., Giannakis O., 2003, poster presented at *Proc. Conf. Multiwavelength Cosmology* (astro-ph/0309628)
- Kodama T., Bell E. F., Bower R. G., 1999, *MNRAS*, 302, 152
- Kurucz R., 1993, *ATLAS9 Stellar Atmosphere Programs and 2 km s<sup>-1</sup> grid*. Kurucz CD-ROM No. 13. Smithsonian Astrophysical Observatory, Cambridge, MA, p. 13
- Le Borgne D., Rocca-Volmerange B., 2002, *A&A*, 386, 446
- Leitherer C., Ferguson H. C., Heckman T. M., Lowenthal J. D., 1995, *ApJ*, 454, L19
- Lin H., Yee H. K. C., Carlberg R. G., Morris S. L., Sawicki M., Patton D. R., Wirth G., Shepherd C. W., 1999, *ApJ*, 518, 533
- Lonsdale C. J. et al., 2003, *PASP*, 115, 897
- McMahon R. G., Walton N. A., Irwin M. J., Lewis J. R., Bunclark P. S., Jones D. H., 2001, *New Astron. Rev.*, 45, 97
- Madau P., 1995, *ApJ*, 441, 18
- Madau P., 1999, in Holt S. S., Smith E. P., eds, *After the Dark Ages: When Galaxies Were Young*. Astron. Soc. Pac., San Francisco, p. 299
- Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, *MNRAS*, 283, 1388
- Massarotti M., Iovino A., Buzzoni A., 2001, *A&A*, 368, 74
- Meurer G. R., Heckman T. M., Calzetti D., 1999, *ApJ*, 521, 64
- Miller G. E., Scalo J. M., 1979, *ApJS*, 41, 513
- Mobasher B., Rowan-Robinson M., Georgakakis A., Eaton N., 1996, *MNRAS*, 282, L7
- Mobasher B. et al., 2004, *ApJ*, 600, L167
- Natali F., Giallongo E., Cristiani S., La Franca F., 1998, *AJ*, 115, 397
- Oliver S. et al., 2000, *MNRAS*, 316, 749
- Pickles A. J., 1998, *PASP*, 110, 863
- Poggianti B. M., Bressan A., Franceschini A., 2001, *ApJ*, 550, 195
- Press W. H., Rybicki G. B., Schneider D. P., 1993, *ApJ*, 414, 64
- Richards G. T. et al., 2001a, *AJ*, 121, 2308
- Richards G. T. et al., 2001b, *AJ*, 122, 1151
- Richards G. T. et al., 2003, *AJ*, 126, 1131
- Rowan-Robinson M., 1995, *MNRAS*, 272, 737
- Rowan-Robinson M., 2003a, *MNRAS*, 344, 13
- Rowan-Robinson M., 2003b, *MNRAS*, 345, 819 (RR03)
- Rowan-Robinson M. et al., 2004, *MNRAS*, 351, 1290
- Savage B. D., Mathis J. S., 1979, *ARA&A*, 17, 73
- Sawicki M. J., Lin H., Yee H. K. C., 1997, *AJ*, 113, 1
- Schneider D. P. et al., 2001, *AJ*, 121, 1232
- Seaton M. J., 1979, *MNRAS*, 187, 73p
- Steidel C. C., Sargent W. L. W., 1987, *ApJ*, 313, 171
- Steidel C. C., Adelberger K. L., Giavalisco M., Dickinson M., Pettini M., 1999, *ApJ*, 519, 1
- Tagliaferri R., Longon G., Andreon S., Capozziello S., Donalek C., Giordano G., 2002, preprint (astro-ph/0203445)
- Thompson R. I., Weymann R. J., Storrie-Lombardi L. J., 2001, *ApJ*, 546, 694
- Vanden Berk D. E. et al., 2001, *AJ*, 122, 549
- Wang Y., Bahcall N., Turner E. L., 1998, *AJ*, 116, 2081
- Webster R. L., Francis P. J., Peterson B. A., Drinkwater M. J., Masci F. J., 1995, *Nat*, 375, 469
- Wolf C. et al., 2001, *A&A*, 365, 681
- Wu X., Zhang W., Zhou X., 2004, *Chinese J. Astron. Astrophys.*, 4, 17
- York D. G. et al., 2000, *AJ*, 120, 1579
- Yoshii Y., Takahara F., 1988, *ApJ*, 326, 1
- Zheng W., Kriss G. A., Telfer R. C., Grimes J. P., Davidsen A. F., 1997, *ApJ*, 475, 469

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.